

N° 92-371-XPB au catalogue

Échantillonnage et pondération

Rapports techniques du recensement de 1996

RECENSEMENT

Statistique Canada · Statistics Canada

Canada

Des données sous plusieurs formes

Statistique Canada diffuse les données sous formes diverses. Outre les publications, des totalisations habituelles et spéciales sont offertes. Les données sont disponibles sur Internet, disque compact, disquette, imprimé d'ordinateur, microfiche et microfilm, et bande magnétique. Des cartes et d'autres documents de référence géographiques sont disponibles pour certaines sortes de données. L'accès direct à des données agrégées est possible par le truchement de CANSIM, la base de données ordinolingue et le système d'extraction de Statistique Canada.

Comment obtenir d'autres renseignements

Toute demande de renseignements au sujet du présent produit ou au sujet de statistiques ou de services connexes doit être adressée à l'un des centres de consultation régionaux de Statistique Canada suivants :

Halifax	(902) 426-5331	Regina	(306) 780-5405
Montréal	(514) 283-5725	Edmonton	(780) 495-3027
Ottawa	(613) 951-8116	Calgary	(403) 292-6717
Toronto	(416) 973-6586	Vancouver	(604) 666-3691
Winnipeg	(204) 983-4020		

Vous pouvez également visiter notre site sur le Web : http://www.statcan.ca

Un service d'appel interurbain sans frais est offert **à tous les utilisateurs qui habitent à l'extérieur des zones de communication locale** des centres de consultation régionaux.

Service national de renseignements	**1 800 263-1136**
Service national d'appareils de télécommunications pour les malentendants	**1 800 363-7629**
Numéro pour commander seulement (Canada et États-Unis)	**1 800 267-6677**
Numéro pour commander par télécopieur (Canada et États-Unis)	**1 877 287-4369**

Renseignements sur les commandes et les abonnements

Les prix ne comprennent pas les taxes de vente

Le produit n° 92-371-XPB au catalogue est offert au coût de 15 $ par copie au Canada. À l'extérieur du Canada, le coût est de 15 $ US par copie.

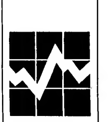
Veuillez commander par la poste, en écrivant à Statistique Canada, Division de la diffusion, Gestion de la circulation, 120, avenue Parkdale, Ottawa (Ontario) K1A 0T6; par téléphone, en composant le **(613) 951-7277** ou le **1 800 700-1033**; par télécopieur, en composant le **(613) 951-1584** ou le **1 800 889-9734**; ou par Internet, en vous rendant à order@statcan.ca. Lorsque vous signalez un changement d'adresse, veuillez nous fournir l'ancienne et la nouvelle adresses. On peut aussi se procurer les produits de Statistique Canada auprès des agents autorisés, dans les librairies et dans les bureaux régionaux de Statistique Canada.

Normes de service à la clientèle

Statistique Canada s'engage à fournir à ses clients des services rapides, fiables et courtois et dans la langue officielle de leur choix. À cet égard, notre organisme s'est doté de normes de service à la clientèle qui doivent être observées par les employés lorsqu'ils offrent des services à la clientèle. Pour obtenir une copie de ces normes de service, veuillez communiquer avec le centre de consultation régional de Statistique Canada le plus près de chez vous.

ISBN 0-660-60571-6
9 780660 605715

92-371-XPB96000



Statistique Canada

Échantillonnage et pondération

Rapports techniques du recensement de 1996

Publication autorisée par le ministre responsable de Statistique Canada

Décembre 1999

N° 92-371-XPB au catalogue

ISBN 0-660-60571-6

Ottawa

Note de reconnaissance

Le succès du système statistique du Canada repose sur un partenariat bien établi entre Statistique Canada et la population, les entreprises, les administrations canadiennes et les autres organismes. Sans cette collaboration et cette bonne volonté, il serait impossible de produire des statistiques précises et actuelles.

Données de catalogage avant publication (Canada)

Vedette principale au titre:

Échantillonnage et pondération

(Rapports techniques du recensement de 1996)
Texte en français et en anglais disposé tête-bêche.
Titre de la p. de t. addit.: Sampling and weighting.
ISBN 0-660-60571-6
CS92-371-XPB

1. Échantillonnage (Statistique).
2. Canada – Recensement, 1996 – Méthodologie.
3. Canada – Recensement, 1996 – Évaluation.
I. Statistique Canada. II. Titre. III. Sampling and Weighting. IV. Collection.

HA741.5 1996 S25 1999 304.6'07'23
C99-988039-XF

Pour citer ce document

Statistique Canada. Échantillonnage et pondération.
Rapports techniques du recensement de 1996.
Ottawa : Industrie Canada, 1999.
Numéro 92-371-XPB au catalogue.

Le papier utilisé dans le présent document répond aux exigences minimales de l'American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials. ANSI Z39.48-1984
♾

Table des matières

Page

Liste des tableaux

I. Introduction

Le recensement de 1996 a nécessité la participation de l'ensemble de la population canadienne, soit environ 29 millions de personnes réparties sur un territoire couvrant 9,2 millions de kilomètres carrés. Une entreprise de cette envergure a constitué un énorme défi à relever. Même si des normes de qualité rigoureuses régissent la collecte et le traitement des données, et malgré les efforts déployés afin de réduire la non-réponse, à l'aide des communications notamment, il est impossible de supprimer toutes les erreurs. Bien que ce terme ne signifie pas nécessairement que des fautes ont été commises, certains éléments d'erreur sont inévitables compte tenu des décisions prises afin de limiter les coûts du recensement.

Statistique Canada doit expliquer les méthodes et les concepts utilisés pour recueillir et traiter les données; il doit en outre renseigner les utilisateurs sur la qualité des données produites, ainsi que sur d'autres caractéristiques des données qui pourraient limiter leur utilité ou leur interprétation. Le présent rapport vise à informer les utilisateurs de la complexité des données et des problèmes susceptibles d'influer sur la façon de les utiliser. Il explique le cadre théorique et les définitions ayant servi à recueillir les données, ainsi que les circonstances inhabituelles pouvant avoir une incidence sur la qualité des données. Par ailleurs, il traite la saisie des données, le contrôle et l'imputation, ainsi que la comparabilité chronologique des données.

Le présent rapport, qui porte sur l'échantillonnage et la pondération, a été préparé par Peter Dick, Karen Switzer, Sylvain Thivierge et Patrick Mason, avec le concours du personnel de deux divisions de Statistique Canada : la Division des méthodes d'enquêtes sociales et la Division des opérations du recensement.

Les utilisateurs trouveront des renseignements supplémentaires sur les concepts, les variables et la géographie du recensement dans le *Dictionnaire du recensement de 1996* (n° 92-351-XPF au catalogue), de même qu'un aperçu des différentes étapes du recensement dans *Le recensement en bref* (n° 92-352-XPF au catalogue).

De nos jours, l'échantillonnage constitue une technique largement utilisée dans maints domaines. Nous pouvons juger de la qualité d'un produit à l'aide d'un échantillon avant d'en faire l'achat; nous formons une opinion des gens à partir d'échantillons de leur comportement; nos impressions sur tel pays ou telle ville sont basées sur les courts voyages que nous y avons faits. Ce sont là des exemples d'échantillonnage où l'on tire des conclusions sur le « tout » à partir de connaissances portant sur une « partie ».

D'une façon plus scientifique, l'échantillonnage est utilisé par les comptables, par exemple, lorsqu'ils effectuent la vérification des états financiers; dans l'industrie, pour vérifier la qualité d'articles produits en série, et par les initiateurs de sondages d'opinion et d'enquêtes pour recueillir des renseignements sur les opinions ou les caractéristiques d'une population donnée. On a généralement recours à l'échantillonnage pour réduire les coûts ou pour obtenir rapidement des données, ou les deux. Dans certains cas, le mesurage peut détruire le produit à l'étude (par exemple, l'évaluation de la durée de vie des ampoules électriques) et par conséquent, l'échantillonnage s'avère essentiel. Le désavantage de cette méthode est que les résultats obtenus à partir d'un échantillon peuvent ne pas être aussi précis que ceux recueillis auprès de la population totale. Toutefois, l'échantillonnage peut être avantageux quand la perte de précision (qui peut être minime si l'échantillon est suffisamment grand) est acceptable en fonction de l'utilisation prévue des résultats. De plus, la réduction de l'envergure d'une étude faite à partir d'un échantillon peut mener à la réduction des erreurs non dues à l'échantillonnage, ce qui compense en partie le manque de précision attribuable à l'échantillonnage.

Pour le recensement de la population de 1996, on a appliqué les méthodes d'échantillonnage de plusieurs façons. On y a eu recours pour s'assurer que la qualité du travail des recenseurs dans le cadre de la collecte des questionnaires respectait certaines normes; on y a eu recours pour le contrôle qualitatif du codage des réponses au cours du dépouillement dans les bureaux régionaux; on y a eu recours pour estimer le sous-dénombrement et le

surdénombrement qui se sont produits pour diverses raisons; on y a eu recours pour évaluer la qualité des données du recensement. Toutefois, la principale application de l'échantillonnage au recensement a été effectuée pendant le dénombrement sur le terrain où toutes les données, excepté les données de base, ont été recueillies auprès d'un échantillon de ménages. Le présent rapport décrit cette dernière utilisation de l'échantillonnage et présente une évaluation de l'incidence de cette méthode sur la qualité des données du recensement.

Le chapitre II fait l'historique de l'application des méthodes d'échantillonnage aux recensements canadiens et décrit celles qui ont été utilisées au recensement de 1996. Le chapitre III décrit les méthodes de pondération des données-échantillon utilisées en vue d'obtenir des estimations à l'échelle de la population totale et explique les considérations d'ordre pratique et théorique qui régissent le choix de ces méthodes. On présente au chapitre IV les études effectuées dans le cadre du programme d'évaluation des méthodes d'échantillonnage et de pondération utilisées au recensement de 1996. Les chapitres V à VIII donnent les résultats de ces études. Le chapitre IX présente certaines conclusions sur les méthodes de pondération utilisées en 1996 et quelques suggestions pour le recensement de 2001.

II. Échantillonnage dans les recensements canadiens

Dans un recensement de la population, l'échantillonnage consiste à recueillir et à traiter certaines caractéristiques à partir d'un échantillon aléatoire de logements et de personnes dénombrés lors du recensement intégral. On obtient ensuite pour l'ensemble de la population les totaux pour ces caractéristiques en pondérant les données-échantillon. Les caractéristiques recueillies pour tous les logements ou toutes les personnes faisant partie du recensement sont appelées « caractéristiques de base », tandis que les caractéristiques recueillies à partir d'un échantillon sont appelées « caractéristiques-échantillon » du recensement.

A. Historique de l'échantillonnage au recensement canadien[1]

L'échantillonnage a été utilisé pour la première fois au cours du recensement canadien de 1941. Un questionnaire sur le logement a été remis à tous les 10 logements de chaque sous-district de recensement. Les réponses aux 27 questions de ce questionnaire ont été intégrées à celles du questionnaire sur la population et les ménages correspondant à ces logements, ce qui a permis de faire le recoupement des caractéristiques-échantillon et des caractéristiques de base. Lors de ce même recensement, on a utilisé l'échantillonnage au cours de l'étape du dépouillement pour obtenir des estimations provisoires du revenu des salariés, de la répartition de la population en âge de travailler et de la composition des familles au Canada. Dans ce cas, l'échantillon était constitué du dixième de tous les secteurs de dénombrement du Canada et tous les questionnaires sur la population de ces secteurs ont été dépouillés à l'avance.

L'échantillonnage a également été utilisé pour le recensement du logement de 1951. Cette fois, on a choisi chaque cinquième logement (ceux dont le numéro d'identification se terminait par 2 ou par 7) et le questionnaire sur le logement comportait 24 questions. Au cours du recensement de 1961, on a prélevé un échantillon de 20 % des ménages privés et on a demandé aux membres de 15 ans et plus de ces ménages de remplir un questionnaire-échantillon sur la population comportant des questions sur la migration interne, la fécondité et le revenu. L'échantillonnage n'a pas été utilisé au cours des recensements de moindre envergure de 1956 et 1966.

Plusieurs innovations importantes ont été apportées aux méthodes de recensement de 1971. La première a été l'utilisation de l'autodénombrement pour la majorité de la population au lieu de la méthode traditionnelle de recensement par représentant. Ce changement découlait des résultats de plusieurs études effectuées au Canada et ailleurs (Fellegi, 1964; Hansen et al., 1959) selon lesquels la présence d'un représentant avait un effet important sur la variance[2] de chiffres du recensement. On prévoyait que l'autodénombrement permettrait de réduire la variance des chiffres du recensement attribuable à la présence du représentant et, en même temps, donnerait au répondant le temps voulu pour remplir le questionnaire à son aise; par conséquent, on s'attendait à ce que les réponses soient plus exactes que par le passé.

Un autre changement survenu dans le recensement de 1971 a été la modification du contenu. Le nombre de sujets traités et de questions posées était plus élevé qu'aux recensements précédents. Une étude des coûts, du fardeau du répondant et des délais de production par rapport à la qualité des données dans le cadre d'un recensement par échantillonnage et par autodénombrement a présidé à la décision de recueillir toutes les données, sauf celles portant

[1] Des renseignements détaillés sur les recensements particuliers sont contenus dans le Rapport administratif, la Revue générale, le Guide sommaire ou Le recensement en bref des recensements en question. Ces produits sont mentionnés en bibliographie à la fin de ce rapport.

[2] La « variance » d'une estimation est une mesure de sa précision. La variance est expliquée plus en détail au chapitre VIII.

sur quelques caractéristiques de base, auprès d'un échantillon d'un tiers de la population au recensement de 1971. Sauf dans les régions les plus isolées du Canada, le tiers des ménages privés a reçu le questionnaire complet comportant toutes les questions du recensement, tandis que le reste des ménages privés a reçu le questionnaire abrégé qui ne contenait que les questions de base soit le nom, le lien de parenté avec le chef de ménage, le sexe, la date de naissance, l'état matrimonial, la langue maternelle, le type de logement, le mode d'occupation, le nombre de pièces, l'alimentation en eau, les toilettes ainsi que certaines questions en rapport avec la couverture du recensement. Tous les ménages vivant dans des régions éloignées prédésignées et dans des logements collectifs[3] ont reçu le questionnaire complet. Le rapport statistique « Sampling in the Census » (Bureau fédéral de la statistique, 1968) donne une description détaillée des projets d'application de cette méthode au recensement de 1971.

Le contenu du recensement de 1976 était beaucoup moindre que celui de 1971. En outre, le questionnaire de 1976 ne contenait pas les questions qui posent le plus de difficulté de collecte, par exemple les questions sur le revenu, ni celles pour lesquelles les frais de codage sont le plus élevés, par exemple celles sur la profession, l'activité économique et le lieu de travail. Par conséquent, les avantages de l'échantillonnage en termes de coûts et d'allégement du fardeau du répondant étaient moins évidents qu'au recensement de 1971. Néanmoins, après avoir évalué les économies éventuelles reliées à divers taux d'échantillonnage et étudié les conséquences sur le plan des relations publiques du retour au dénombrement intégral (l'échantillonnage ayant fait ses preuves en 1971), il a été décidé d'appliquer en 1976 la même méthode d'échantillonnage qu'en 1971.

Au recensement de 1981, on a utilisé la plupart des mêmes méthodes qu'en 1971 et 1976, sauf que le taux d'échantillonnage est passé du tiers au cinquième des ménages. Les études réalisées à cette époque ont révélé que la diminution consécutive de la qualité des données (mesurée au moyen de la variance) serait acceptable et qu'elle ne serait pas suffisamment importante pour annuler les avantages découlant de la réduction des coûts et du fardeau de réponse et diminuerait les délais de production (voir Royce, 1983). Douze questions ont été posées à l'ensemble de la population et 34 questions de plus ont été posées aux membres de l'échantillon.

Le recensement de 1986 a été le premier recensement majeur réalisé au milieu d'une décennie. Il avait été décidé que seul un recensement majeur pourrait répondre aux besoins croissants en matière de données locales sur le marché du travail, besoins rendus plus pressants par l'importante récession (1981-1982) survenue depuis le recensement précédent. Toutefois, afin de maintenir les coûts d'élaboration au niveau le plus bas possible, on a décidé de faire très peu de changements. Par conséquent, les questions ainsi que les méthodes de collecte et de dépouillement des données du recensement de 1981 ont été reprises, sauf s'il y avait des raisons convaincantes de ne pas procéder ainsi. Des questions portant sur huit sujets posées au recensement de 1981 n'ont pas été reprises en 1986, tandis que trois nouvelles questions ont été ajoutées.

En 1991, le recensement de la population englobait à la fois les résidents permanents et, pour la première fois, les résidents non permanents – personnes qui sont titulaires d'un permis de séjour pour étudiants, d'un permis de travail ou d'un permis ministériel, ou qui revendiquent le statut de réfugié. Afin qu'on puisse savoir qui étaient les résidents non permanents, une nouvelle question a dû être élaborée, testée et incluse dans le recensement de 1991. En tout, 12 nouvelles questions ont été ajoutées au recensement de 1991, tandis que quatre questions du recensement de 1986 n'ont pas été posées en 1991. Une fois terminée la collecte des questionnaires du recensement de 1991, deux enquêtes postcensitaires ont été effectuées. Une de ces enquêtes, l'Enquête sur la santé et les limitations d'activités (ESLA), visait à recueillir des données sur la santé et le bien-être des Canadiens en général. La deuxième enquête, soit celle auprès des peuples autochtones, visait à recueillir des renseignements sur la population autochtone vivant dans les réserves et hors des réserves. De plus, en 1991, on a noté une augmentation importante de l'automatisation du traitement des données et des véhicules par lesquels les produits et services sont élaborés et livrés au client.

[3] Un logement collectif est une institution ou un établissement commercial ou communautaire, par exemple un hôtel, un hôpital, une résidence de personnel ou un camp de chantier.

B. Plan d'échantillonnage utilisé lors du recensement de 1996

Le jour du recensement en 1996, une multitude de renseignements ont été recueillis auprès de chaque Canadien. Il faut noter qu'en 1996, le jour du recensement a été déplacé de la date traditionnelle du début juin pour être fixée au 14 mai 1996. La majeure partie de ces informations ont été obtenues à partir d'un échantillon. Dans chaque secteur d'autodénombrement, un échantillon de 1 sur 5 des logements privés occupés a été choisi pour recevoir un questionnaire complet (formule 2B) contenant toutes les questions du recensement. Les questions de base sur l'âge, le sexe, l'état matrimonial, la langue maternelle et le lien avec la personne repère du ménage (Personne 1) ont été posées à toutes les personnes alors que les questions additionnelles portant sur le genre de logement, le mode d'occupation et les questions socioéconomiques ont été posées à un échantillon de 20 % de la population.

Tous les logements se trouvant dans les secteurs dénombrés par un recenseur (généralement les régions éloignées ou les réserves indiennes) ont reçu le questionnaire 2B. Tous les logements collectifs ont également reçu la formule 2B. Par contre, les personnes suivantes dans les logements collectifs n'ont pas eu à répondre aux questions posées à l'échantillon de la population :

a) les détenus dans les établissements de correction et les établissements pénitentiaires ou les prisons;

b) les patients dans les hôpitaux généraux, les établissements de soins spéciaux, les établissements pour les personnes âgées ainsi que les établissements pour personnes souffrant de maladies chroniques et les hôpitaux psychiatriques;

c) les foyers collectifs pour enfants (orphelinats) et les foyers pour jeunes contrevenants.

Les Canadiens en poste à l'étranger (en général, du personnel d'ambassade ou des membres des Forces armées) ont reçu une formule 2C, qui contenait les mêmes questions que la formule 2B mais qui ne comportait pas de questions sur le logement. Les répondants devaient toutefois indiquer leur lieu habituel de résidence au Canada. Les renseignements sur les logements privés inoccupés ont été consignés sur une formule 2A.

D'après les procédures de livraison de base, le recenseur devait planifier un itinéraire couvrant tous les logements de son secteur de dénombrement (SD) puis se rendre à chaque logement et y laisser un questionnaire. Le choix de l'échantillon, c'est-à-dire la décision quant au genre de questionnaire à livrer à chaque logement occupé, a été fait à partir du Registre des visites (RV) dans lequel le recenseur inscrivait tous les logements de son secteur. Chaque cinquième ligne du registre était ombrée, indiquant qu'un questionnaire 2B devait être laissé. Un point de départ aléatoire était déterminé à partir de la première ligne du RV, et le recenseur sautait zéro, une, deux, trois ou quatre lignes selon que le cinquième, quatrième, troisième, deuxième ou premier logement du SD devait être le premier à recevoir le questionnaire complet. Par la suite, chaque logement inscrit sur une ligne ombrée recevait automatiquement le questionnaire complet. Ces procédures étaient expliquées dans le manuel du recenseur et une attention particulière y était portée pendant la formation du recenseur afin de réduire au minimum le risque d'erreur dans la sélection de l'échantillon. Des vérifications du travail des recenseurs, à des fins de contrôle qualitatif, ont été effectuées par le commissaire au recensement.

Pour le recensement de 1996, un essai important sur les opérations du recensement a été réalisé par la poste auprès de plus de 400 000 ménages dans l'est de l'Ontario, y compris la région d'Ottawa. Tous les ménages des régions urbaines ont été inscrits dans un registre des adresses puis des questionnaires ont été expédiés à tous ces logements. Au préalable, les ménages-échantillon (2B) avaient été sélectionnés systématiquement à partir de ce registre. Les répondants étaient invités à retourner leur questionnaire par la poste; un petit groupe de personnes très qualifiées ont assuré un suivi auprès des non-répondants.

Dans la terminologie de l'échantillonnage, l'échantillon du recensement peut être décrit comme un échantillon stratifié avec sélection systématique de logements privés occupés, choisi en fonction d'un taux d'échantillonnage constant de 1 sur 5 dans toutes les strates (SD). En tant qu'échantillon de personnes, il peut être considéré comme un échantillon stratifié avec sélection systématique de grappes, un logement constituant une grappe. Sarndal, Swensson et Wretman , 1992, donnent une description détaillée des concepts et de la terminologie se rapportant à l'échantillonnage.

C. Traitement des données de l'échantillon du recensement

Dès que le recenseur était en possession du questionnaire rempli (2A ou 2B) de chaque logement de son secteur et que son travail avait été approuvé, les questionnaires étaient envoyés à l'un des sept centres de dépouillement régionaux aux fins du traitement manuel. Les données complètes pour chaque SD ont été saisies et stockées sur bandes magnétiques. Les questionnaires et les bandes magnétiques ont ensuite été envoyés au bureau central à Ottawa. Après réception au bureau central, des contrôles automatisés étaient effectués aux fins d'incohérences dans les données, ces incohérences devant être résolues manuellement. Une fois terminées toutes les mises à jour des données pour un SD, les données étaient restructurées et transférées à la section Contrôle et Imputation.

Les données étaient chargées dans les bases de données du contrôle et de l'imputation, organisées par données 2A (100 %) et données 2B (20 %), en cinq régions pour chacune des bases. Les bases de données 2A contenaient les caractéristiques démographiques de base pour l'ensemble de la population et les bases de données 2B contenaient les données-échantillon (20 %). Les données ont été soumises à une série de modules de traitement spécialisés, où tous les problèmes liés aux données invalides, incohérentes ou manquantes ont été résolus. Les bases de données 2A ont été traitées en premier et ont donné lieu à la création de la Base de données d'extraction 2A définitive pour le Canada.

Une fois la dernière main mise aux données intégrales, les données-échantillon (20 %) ont été traitées. Les enregistrements de non-réponse 2B ont été supprimés de la base de données 2B. Une base de données d'extraction 2B définitive a été créée. Cette base de données contenait à la fois les données intégrales et les données-échantillon uniquement pour les ménages et les personnes de l'échantillon. Les coefficients de pondération calculés à l'aide des données intégrales (selon la méthode décrite au chapitre III) ont été incluses dans cette base de données.

III. Estimations basées sur l'échantillon du recensement

Tout échantillonnage nécessite l'application d'une méthode d'estimation pour rapporter les données-échantillon à l'échelle de la population totale. La méthode d'estimation est généralement choisie en fonction d'exigences d'ordre opérationnel et théorique. Sur le plan opérationnel, cette méthode doit être compatible avec le système de dépouillement auquel elle s'intègre, tandis que sur le plan théorique, elle doit minimiser l'erreur d'échantillonnage des estimations produites. Les deux sections qui suivent décrivent les considérations opérationnelles et théoriques qui régissent le choix des méthodes d'estimation à partir de l'échantillon du recensement.

A. Considérations d'ordre opérationnel

Une méthode d'estimation peut être décrite mathématiquement à l'aide d'une formule algébrique qui indique comment la valeur de l'estimateur de la population est calculée en fonction des valeurs observées dans l'échantillon. Quand il s'agit de petites enquêtes qui ne portent que sur une ou deux caractéristiques, ou lorsque la formule d'estimation est très simple, il est possible de calculer les estimations-échantillon en appliquant cette formule aux données-échantillon pour chaque estimation requise. Toutefois, dans le cas d'une enquête ou d'un recensement portant sur une gamme étendue de caractéristiques, ou lorsque la formule d'estimation est assez complexe, il n'est pas possible d'appliquer une formule séparément pour chaque estimation désirée. Dans le cas du recensement, par exemple, chaque cellule de chaque totalisation, fondée sur des données-échantillon recueillies à chaque niveau géographique, représente une estimation-échantillon qui nécessiterait, sous cette approche, une application distincte de la formule d'estimation. Par ailleurs, les diverses estimations effectuées par calcul individuel à partir d'un même échantillon du recensement ne sont pas nécessairement cohérentes.

Par conséquent, la méthode utilisée pour le recensement et pour de nombreuses enquêtes-échantillon consiste à décomposer le processus d'estimation en deux étapes : a) le calcul des coefficients de pondération, c'est-à-dire la pondération elle-même et b) la sommation des coefficients de pondération afin de produire des estimations des chiffres de population. Les seules difficultés mathématiques, s'il y en a, se rapportent à l'étape a) qui n'est effectuée qu'une fois, tandis que l'étape b) se limite à un simple processus de sommation des coefficients de pondération au moment où une totalisation est extraite. Il faut noter que, comme le coefficient de pondération relié à une unité de l'échantillon est le même quelle que soit la totalisation effectuée, la cohérence entre les diverses estimations basées sur les données-échantillon est assurée.

B. Considérations d'ordre théorique

Selon la théorie de l'échantillonnage, il est possible de déterminer, à partir d'un plan d'échantillonnage particulier et d'une procédure d'estimation donnée, les chances qu'un certain intervalle contienne la valeur (inconnue) que l'on veut estimer pour la population entière. Le principal objet de l'estimation est de réduire au maximum l'étendue de ces intervalles de façon que les hypothèses sur les valeurs inconnues de la population soient aussi précises que possible. La mesure courante de précision utilisée pour comparer les procédures d'estimation est appelée l'erreur-type. Pourvu que certaines conditions relativement peu contraignantes soient respectées, un intervalle de plus ou moins deux erreurs-types par rapport à la valeur estimée renfermera la valeur de la population pour environ 95 % de tous les échantillons possibles.

Outre la réduction au minimum de l'erreur-type, le deuxième objectif relatif au choix d'une méthode d'estimation de l'échantillon du recensement est de garantir autant que possible que les estimations faites à partir de l'échantillon des caractéristiques de base (c.-à-d. 2A) soient cohérentes avec les valeurs connues de la population correspondante.

Heureusement, ces deux objectifs sont généralement complémentaires, c'est-à-dire que l'erreur d'échantillonnage est habituellement réduite lorsqu'on s'assure que les estimations-échantillon de certaines caractéristiques de base sont compatibles avec les chiffres correspondants de la population. Notons cependant qu'il peut arriver que le fait de forcer les estimations-échantillon de caractéristiques de base à être compatibles avec les chiffres correspondants de la population pour de très petits sous-groupes puisse avoir des conséquences défavorables sur l'erreur-type des estimations des caractéristiques-échantillon proprement dites.

Lorsqu'on n'a aucune information sur la population échantillonnée autre que celle qui a été recueillie pour les unités d'échantillonnage, la méthode d'estimation se limite à pondérer les unités d'échantillonnage en proportion inverse de leur probabilité de sélection; par exemple, si toutes les unités avaient une chance sur cinq d'être choisies, toutes les unités recevraient un coefficient de pondération de 5. En pratique, toutefois, on a généralement une certaine connaissance de la population observée, par exemple, sa taille globale et, peut-être, sa répartition en fonction d'une variable donnée (disons, par province). Ces données peuvent être utilisées pour améliorer la formule d'estimation de façon à produire des estimations qui se rapprochent davantage de la valeur inconnue de la population. Dans le cas de l'échantillon du recensement, on dispose d'un grand nombre de données très détaillées sur la population échantillonnée, soit les caractéristiques de base obtenues pour tous les ménages, à tous les niveaux géographiques. On peut se servir de cette multitude d'informations sur la population pour améliorer les estimations faites à partir de l'échantillon du recensement mais, d'autre part, ces données peuvent également occasionner des ennuis compte tenu du fait **qu'il est impossible d'obtenir des estimations-échantillon des caractéristiques de base qui soient compatibles avec toutes les données sur la population à tous les niveaux géographiques**. Les écarts entre les estimations-échantillon et les valeurs de la population apparaissent quand on produit un recoupement d'une variable de base et d'une variable-échantillon. La totalisation doit être fondée sur les données-échantillon; de ce fait, les totaux marginaux des variables de base sont des estimations-échantillon que l'on peut comparer avec les chiffres correspondants de la population d'une autre totalisation basée sur les données intégrales. Il se peut que cette comparaison ne donne pas une concordance exacte.

C. Élaboration d'une méthode d'estimation pour l'échantillon du recensement

Étant donné qu'il faut attribuer un coefficient de pondération à chaque unité d'échantillonnage (personne, famille ou ménage), le plus simple serait de choisir le coefficient 5, puisqu'il s'agit d'un échantillon de 1 sur 5. Une telle méthode serait simple et sans biais[4], et si l'on n'avait pas d'autres données que celles de l'échantillon, cela pourrait être la meilleure méthode. Or, bien qu'on sache que l'échantillon contiendra presque exactement le cinquième de tous les ménages (à l'exclusion des ménages collectifs et de ceux qui se trouvent dans les secteurs de recensement par interview), on ne peut pas être certain qu'il contiendra exactement un cinquième de toutes les personnes, ou un cinquième de chaque genre de ménage, ou un cinquième de toutes les femmes de 25 à 34 ans, etc. Par conséquent, une telle méthode n'assurerait pas la cohérence, même pour les sous-groupes les plus importants de la population. Dans ce dernier cas, les fractions seraient très proches d'un cinquième, mais elles pourraient s'en écarter de façon sensible dans le cas des petits sous-groupes. Une autre méthode simple serait de prendre certains sous-groupes importants, par exemple les groupes d'âge-sexe par province, puis, pour chacun, compter le nombre d'éléments de la population appartenant au sous-groupe (N) et le nombre d'éléments appartenant à l'échantillon (n), et attribuer à chaque unité d'échantillonnage appartenant au sous-groupe un coefficient de pondération égal à N/n.

---

[4] « Sans biais » veut dire que la moyenne des estimations de l'ensemble de tous les échantillons possibles serait égale à la valeur vraie de la population.

Par exemple, si l'on recensait 5 000 hommes de 20 à 24 ans à l'Île-du-Prince-Édouard, et si 1 020 de ces hommes faisaient partie des ménages inclus dans l'échantillon, un coefficient de pondération de 5 000/1 020 = 4,90 serait attribué à chaque homme de 20 à 24 ans dans l'échantillon de l'Île-du-Prince-Édouard. De cette façon, chaque fois que les groupes d'âge-sexe de cinq ans seraient recoupés avec une caractéristique-échantillon pour l'Île-du-Prince-Édouard, le total marginal pour le groupe d'âge-sexe d'hommes de 20 à 24 ans concorderait avec le total de la population, soit 5 000. Ce genre de méthode d'estimation est appelée « estimation par la méthode du quotient ». Il est à noter que, dans cet exemple particulier, un coefficient de pondération de 5 donnerait une estimation-échantillon de 5 100 (1 020 x 5). La méthode d'estimation qui a été utilisée avant le recensement de 1991 était une généralisation de l'estimation par quotient appelée méthode itérative du quotient (MIQ). Pour plus de détails sur la MIQ, voir le *Guide à l'intention des utilisateurs sur la qualité des données du recensement de 1986 : Échantillonnage et pondération* (Statistique Canada, 1990), ainsi que Brackstone et Rao, 1979.

Pour le recensement de 1991, on a décidé de recourir à une autre méthode d'estimation, appelée procédure d'estimation généralisée des moindres carrés (PEGMC) en deux étapes. Ce choix visait à améliorer la concordance entre les chiffres de population et les estimations correspondantes au niveau des SD, par rapport à ce que permettait la MIQ. Les erreurs-types des estimations produites par la PEGMC pour les petites régions géographiques étaient également réduites. En outre, la PEGMC permettait de déterminer, pour chaque ménage de l'échantillon, un coefficient de pondération unique pouvant servir à produire des estimations tant pour les caractéristiques des personnes que pour les caractéristiques des ménages. Par contre, avec la MIQ, il fallait utiliser des coefficients de pondération différents pour produire des estimations des caractéristiques des ménages et des personnes, ce qui entraînait parfois des incohérences.

Avec la PEGMC, les coefficients de pondération d'environ 5 ont été ajustés aussi peu que possible pour chaque ménage, de sorte qu'on s'assurait qu'il y avait une concordance parfaite entre les estimations et les chiffres de population pour autant de caractéristiques de base que possible qui sont énoncées à l'annexe B (celles-ci seront appelées contraintes). Il fallait que cette concordance parfaite soit réalisée au niveau des régions de pondération (RP). Chaque RP contenait, en moyenne, sept SD échantillonnés. On retrouve plus d'information sur les RP au chapitre VI, section A du présent rapport. On y explique que la PEGMC est un estimateur par régression. L'ouvrage de Sarndal, Swensson et Wretman, 1992, qui renferme plus de détails sur les estimateurs par régression, constitue un bon outil de référence.

D. Procédure d'estimation généralisée des moindres carrés (PEGMC) en deux étapes

Les calculs des coefficients de pondération sont faits indépendamment dans chaque région de pondération (RP). Certaines des contraintes énoncées à l'annexe B (tant au niveau des SD qu'à celui des RP) doivent être retranchées pour chaque RP, de sorte que la concordance entre les chiffres de population et les estimations ne peut être garantie pour toutes les contraintes. Les contraintes sont initialement retranchées au niveau des RP pour les raisons suivantes :

- elles s'appliquent à moins de 20 ménages (on les appelle des contraintes **faibles**);
- elles sont redondantes (on les appelle des contraintes **linéairement dépendantes** [LD]);
- elles sont quasi redondantes (on les appelle des contraintes **quasi linéairement dépendantes** [QLD]).

Une contrainte redondante serait, par exemple, le nombre total de femmes, puisque les contraintes sont déjà présentes lorsque le nombre total de personnes et le nombre total d'hommes sont utilisés. On est en présence d'une **contrainte linéairement dépendante** lorsque deux contraintes (ou plus) font en sorte que la troisième sera automatiquement satisfaite. Pour illustrer ce qu'on entend par **contrainte quasi redondante**, prenons le cas des contraintes représentant les personnes dont **l'état matrimonial = célibataire** et **taille du ménage = 1**. Si la plupart, mais non la

totalité des personnes dans les ménages de taille 1 sont célibataires, les deux contraintes sont donc presque équivalentes, et l'on peut en considérer une comme QLD. Les contraintes LD ont été retranchées pour accroître l'efficacité de calcul de l'algorithme de pondération. Les contraintes faibles et les contraintes QLD ont été retranchées parce que, si on ne l'avait pas fait, les estimations auraient pu devenir instables et présenter des erreurs-types élevées.

Une fois retranchées les contraintes faibles, LD et QLD au niveau des RP, le calcul des coefficients de pondération PEGMC a lieu en deux étapes. À la première étape, les coefficients de pondération initiaux, qui égalent l'inverse de la fraction d'échantillonnage des ménages dans les SD, sont ajustés individuellement pour chaque SD. Il faut noter que certaines contraintes, qui n'ont pas été retranchées au niveau des RP, peuvent l'être au niveau des SD, parce qu'elles sont faibles ou linéairement dépendantes. Les contraintes qui demeurent pour le SD sont triées selon le nombre de ménages auxquels elles s'appliquent au niveau du SD. Elles sont ensuite divisées en deux groupes, soit ceux de nombre pair et ceux de nombre impair. Les coefficients de pondération PEGMC sont calculés au niveau du SD pour chaque groupe de contraintes. Parfois, la méthode d'estimation donne des coefficients de pondération très faibles (inférieurs à 10) ou très élevés (supérieur à 25) pour obtenir la concordance nécessaire à l'égard de certaines contraintes. Ces coefficients, appelés coefficients de « valeur aberrante », ne sont pas désirables. Par conséquent, quand cela se produit, les contraintes qui en sont responsables sont repérées et retranchées, et les coefficients sont calculés de nouveau. Enfin, on fait la moyenne des coefficients obtenus pour les deux groupes de contraintes pour chaque ménage de l'échantillon, ce qui donne les coefficients de pondération de la première étape pour chaque SD.

Les coefficients de pondération produits à la première étape sont utilisés comme coefficients initiaux de la deuxième étape, qui consiste à les ajuster de façon à obtenir une concordance entre les estimations-échantillon et les chiffres de population au niveau des RP. Toutes les contraintes non désignées comme faibles, LD or QLD au niveau de la RP sont utilisées. Ici encore, si des valeurs aberrantes sont produites, les contraintes responsables sont repérées et retranchées, et les coefficients définitifs sont calculés de nouveau. Même si la deuxième étape détruit quelque peu la concordance obtenue pour les estimations au niveau des SD à la première étape, les estimations finales au niveau des SD sont tout de même plus proches des chiffres de population que si la première étape avait été omise. Pour une explication plus détaillée du calcul des coefficients de pondération, voir Bankier, Rathwell et Majkowski, 1992.

En 1996, deux estimations distinctes ont été effectuées. La deuxième estimation a été jugée nécessaire lorsque certaines divergences inacceptables ont été observées pour les variables « union libre » et « taille du ménage » lors de la première estimation. Pour la deuxième estimation, une nouvelle contrainte (union libre = oui) a été ajoutée et le coefficient de pondération initial a été modifié. Lors de la première estimation, le coefficient de pondération initial des ménages correspondait à l'inverse de la fraction d'échantillonnage des ménages. Cependant, ces coefficients de pondération initiaux ont été ajustés lors de la deuxième estimation pour obtenir une concordance entre les estimations relatives à la taille des ménages et les chiffres de population au niveau des RP. Pour obtenir de plus amples renseignements, consulter Bankier, Houle et Luc, 1997.

Des coefficients de pondération ont été établis à l'aide de la PEGMC uniquement pour les ménages privés ayant reçu le questionnaire complet du recensement parmi un SD échantillonné (1 sur 5). Un coefficient de pondération de 0 a été attribué aux ménages privés du SD échantillonné qui avaient reçu une formule abrégée. On a attribué un coefficient de pondération 1 à tous les ménages privés des SD non échantillonnés, car la totalité des répondants de ces secteurs ont fourni des renseignements sur la formule 2B. Les ménages collectifs ont aussi reçu un coefficient de pondération de 1.

IV. Programme d'évaluation de l'échantillonnage et de la pondération

Le programme d'évaluation de l'échantillonnage et de la pondération avait pour objet de déterminer l'effet de l'échantillonnage et de la pondération sur la qualité des données-échantillon du recensement. À cette fin, quatre études ont été réalisées afin de mesurer la qualité des données-échantillon et des estimations-échantillon du recensement ainsi que de fournir des renseignements utiles à la planification des futurs recensements. Les études ont porté sur les sujets suivants :

a) examen du biais d'échantillonnage;

b) évaluation des méthodes de pondération;

c) évaluation de la cohérence des estimations-échantillon et des chiffres de population;

d) étude visant à évaluer la variance d'échantillonnage pour diverses caractéristiques-échantillon (20 %).

La suite du présent chapitre décrit brièvement ces quatres études. Les chapitres V à VIII en présentent les résultats.

A. Examen du biais d'échantillonnage

Dans toute enquête, un biais entachant les réponses peut provenir de diverses sources. Le but de cette étude était de déterminer si les réponses aux questions de base du questionnaire 2B étaient biaisées de quelque façon que ce soit et de cerner, si possible, les causes de tout biais observé.

B. Évaluation des méthodes de pondération

Le but de cette étude était d'évaluer l'efficacité de la PEGMC. À cette fin, on a examiné le niveau de concordance entre les estimations-échantillon et les chiffres de population à l'égard des diverses contraintes, pour l'ensemble des RP au Canada. Le nombre et le type des contraintes retranchées au niveau des RP, ainsi que les raisons pour lesquelles elles avaient été retranchées, ont été étudiés pour expliquer les incohérences observées. En outre, on a examiné la distribution des coefficients de pondération PEGMC, ainsi que les écarts entre les résultats de 1996 et ceux de 1991.

C. Examen de la cohérence des estimations-échantillon et des chiffres de population

Dans cette étude, on a examiné le niveau de concordance, ou de cohérence, entre les estimations-échantillon et les chiffres de population pour un large éventail de caractéristiques de base. On a étudié cette concordance pour diverses unités géographiques autres que les RP. On a également comparé la cohérence obtenue en 1996 à celle obtenue en 1991 pour ces caractéristiques.

D. Variance d'échantillonnage

La « variance » d'une estimation est une mesure de sa précision. Il est relativement peu coûteux de calculer des estimations de la variance pour des estimateurs lorsqu'on utilise des coefficients de pondération égaux à 5 et qu'on suppose un échantillonnage aléatoire simple. Toutefois, il est beaucoup plus coûteux d'estimer la variance des estimateurs du recensement pour le plan d'échantillonnage et les techniques d'estimation utilisés. Des **« facteurs d'ajustement »** ont été calculés pour le recensement de 1986. Les facteurs d'ajustement sont le rapport entre les estimations des erreurs-types (les racines carrées de la variance) des estimations du recensement et les estimations simples des erreurs-types. On peut alors obtenir une estimation de l'erreur-type d'une estimation du recensement pour n'importe quelle caractéristique dans toute unité géographique en multipliant l'estimation simple de l'erreur-type par le facteur d'ajustement approprié. On explique ensuite pourquoi ces estimations de l'erreur-type ne sont pas nécessairement exactes, en raison du biais introduit dans le processus par l'échantillonnage, le traitement des données et la méthode d'estimation.

V. Biais d'échantillonnage

Les estimations établies à partir d'une enquête-échantillon sont sujettes à des erreurs d'échantillonnage. Un type d'erreur d'échantillonnage est causé par la variabilité des caractéristiques de la population. Cette variabilité provient d'échantillons différents qui produisent des estimations différentes, aucune d'elles n'égalant nécessairement la vraie valeur de la population. Toutefois, les estimations égaleront la vraie valeur de la population en moyenne **pourvu que l'échantillon ne comporte pas de biais**. La présence de biais donne lieu à une tendance à la surestimation ou à la sous-estimation de la vraie valeur de la population. Malheureusement, il est souvent difficile d'éliminer le biais complètement. Dans le cas du recensement de la population, les réponses peuvent être entachées d'un biais causé, entre autres, par les erreurs de couverture, le biais de non-réponse, le biais de réponse (p. ex., des recensés qui répondent différemment sur la formule 2B que sur la formule 2A), les erreurs commises par les recenseurs (p. ex., des recenseurs ne choisissant pas l'échantillon selon les exigences établies), les erreurs de traitement, et ainsi de suite.

Le recensement de la population constitue une occasion unique d'examiner avec attention le biais associé aux caractéristiques de base. Il existe essentiellement deux estimations des mêmes caractéristiques de base : les chiffres pour la population totale et les estimations-échantillon (1 sur 5). Le but de l'Examen du biais d'échantillonnage était de déterminer si les réponses aux questions de base de la formule 2B comportaient un biais. On a obtenu les estimations-échantillon en multipliant les chiffres de l'échantillon au niveau du SD par des coefficients de pondération simples égaux à la fraction de sondage inverse des ménages du SD (environ 5) et en faisant ensuite la somme au niveau de la région géographique appropriée[5].

Au départ, les résultats ont été calculés à l'échelle nationale. Le tableau 5.1 montre les résultats obtenus au niveau national pour les chiffres de population et les estimations ainsi que la différence entre ces résultats, l'écart et l'erreur-type des estimations établies pour 32 caractéristiques de base. La valeur Z (qui constitue un test de signification statistique) figure également au tableau; ceux-ci sont expliqués plus en détail ci-dessous. Dans le cas de 22 caractéristiques, la valeur Z se situe à l'extérieur des intervalles prévus, à savoir -2 à 2. Ainsi, les valeurs Z associées aux hommes, aux personnes de 20 à 29 ans, aux personnes de 75 ans et plus et aux hommes de plus de 15 ans sont inférieures à -2, ce qui indique que ces groupes sont sous-représentés dans l'échantillon, au niveau national, par rapport à l'ensemble de la population. Cependant, les valeurs Z associées aux femmes, à la population totale, aux personnes de 0 à 14 ans et aux personnes mariées sont supérieures à 2, ce qui révèle que ces caractéristiques sont surreprésentées dans l'échantillon. En outre, les ménages d'une personne et les ménages de six personnes ou plus étaient sous-représentés, tandis que les ménages de 2 à 5 personnes étaient surreprésentés. En général, l'échantillon contient un trop grand nombre d'enfants, de personnes d'âge moyen, de femmes et de personnes mariées, tandis qu'il ne renferme pas suffisamment de jeunes adultes, de personnes âgées, d'hommes et de personnes non mariées.

Ces différences, qui étaient statistiquement significatives au seuil de confiance de 5 % dans le tableau 5.1, ont été établies à l'aide de la formule suivante :

$$Z^{(0)} = \frac{\hat{X}^{(0)} - X}{\sqrt{V(\hat{X}^{(0)})}} \qquad (1)$$

[5] Ces estimations simples ont été utilisées au lieu des estimations PEGMC, parce que la PEGMC tend à masquer le biais d'échantillonnage en forçant les estimations de caractéristiques de base à égaler les chiffres de population.

Cette formule est expliquée plus en détail à l'annexe C. La statistique obtenue a été calculée à l'aide de données comportant des imputations pour chacune des 281 divisions de recensement (DR) du Canada et pour l'ensemble des 32 caractéristiques où $\hat{X}^{(0)}$ est une estimation fondée sur l'échantillon 2B établie au moyen de coefficients de pondération simples, tandis que X est le chiffre de population connu (2A) et que $V(\hat{X}^{(0)})$ représente la variance d'échantillonnage de l'estimateur $\hat{X}^{(0)}$. Les valeurs $Z^{(0)}$, pour les 281 DR, devraient suivre une distribution à peu près normale (moyenne = 0; variance = 1) si un échantillon aléatoire simple – ou un échantillon systématique – de ménages était sélectionné de façon non biaisée dans chacun des SD et s'il n'était pas touché par le traitement.

Le tableau 5.1 présente les résultats à l'échelle nationale, mais il ne permet pas de déterminer si quelques résultats présentant un biais négatif compensent d'autres résultats présentant un biais positif. Pour trouver réponse à cette question, on présente au tableau 5.2 un résumé des statistiques relatives au biais pour les 281 DR. Le tableau montre la moyenne, l'erreur-type et le résultat du test effectué (qui consistait à déterminer si la moyenne réelle des valeurs Z était de 0) pour les valeurs Z. Cette hypothèse était rejetée si /T/ des valeurs Z était inférieur à 1,96 pour les 32 caractéristiques de base. En règle générale, les caractéristiques pour lesquelles des résultats biaisés avaient été enregistrés (tableau 5.1) à l'échelle nationale donnaient également des résultats biaisés si l'on examinait la totalité des DR. Toutefois, certaines différences sont dignes de mention. Ainsi, au niveau national, les hommes semblaient sous-représentés dans l'échantillon, mais ce biais n'a pas été observé lors de l'examen des DR. Comme, on **présume**, dans le tableau 5.2, **que toutes les DR sont de la même taille**, nous pouvons en déduire que l'estimation des hommes dans les grandes DR (urbaines) présente un biais négatif, tandis que les hommes des petites DR (rurales) sont probablement surreprésentés. En revanche, un biais a été observé à l'échelle nationale en ce qui concerne les ménages de cinq personnes, mais l'examen des DR a révélé qu'ils étaient surreprésentés dans l'échantillon. Ce résultat s'explique par le fait que, même si très peu de DR présentent une sous-représentation ou une surreprésentation élevée, on observe une faible, mais constante, surreprésentation de ce groupe dans tout l'échantillon.

Le tableau 5.2 montre également le pourcentage de DR correspondant des valeurs Z inférieures à -2, supérieures à 2 et supérieures à 0. Si l'échantillon n'était pas biaisé, nous nous attendrions, d'une part, à ce qu'environ 2,5 % des DR affichent des valeurs Z inférieures à -2 ou supérieures à 2 et, d'autre part, à ce que 50 % des DR correspondent à des valeurs Z supérieures à 0. De toute évidence, certaines caractéristiques prouvent que le biais est répandu parmi presque toutes les DR. Par exemple, les personnes mariées sont surreprésentées dans 82 % des DR. Par ailleurs, dans 13 % des DR, les valeurs Z associées aux personnes mariées sont supérieures à 2, alors qu'elles sont inférieures à -2 dans seulement 0,4 % des DR, soit 1 sur 281. L'échantillon du recensement contient manifestement un trop grand nombre de personnes mariées. Par contre, le nombre de ménages d'une personne était surreprésenté (valeur Z supérieure à 0) dans seulement 26 % des DR. Mentionnons également qu'aucune DR n'a obtenu une valeur Z supérieure à 2 pour les ménages d'une personne.

Comme on l'a mentionné dans l'Introduction, les différences observées entre les estimations-échantillon (basées sur une pondération simple) et les chiffres de population peuvent être expliquées de différentes façons. Une cause possible est le fait qu'au recensement de 1996, il n'y a eu que 86 183 ménages (0,8 % du total) qui ont été manquants ou qui ont refusé de répondre aux questions, en comparaison avec 253 156 (2,6 % du total) au recensement de 1991. Ces ménages ont totalement refusé de répondre aux questions ou le recenseur a été incapable d'obtenir des renseignements parce que les membres du ménage étaient absents au moment où le recensement a été mené ou avaient déménagé, le jour même du recensement ou après cette date, sans répondre au questionnaire. Le recenseur pouvait parfois déterminer le nombre de personnes, mais en général toutes les autres réponses devaient être imputées pour ces ménages. De l'ensemble des ménages manquants ou ayant refusé de répondre en 1996, 32 820 (38 % des 86 183 ménages) étaient des ménages de l'échantillon, alors que nous nous serions attendus à ce que 20 % en fassent partie si les taux de non-réponse des ménages 2A et 2B étaient équivalents. En outre, 3 358 ménages de l'échantillon, même s'ils n'étaient pas manquants ou n'avaient pas refusé de répondre (c'est-à-dire qu'ils avaient fourni certaines réponses aux questions de base), n'ont fourni aucune réponse aux questions-échantillon. Au cours du traitement des données, ces 38 820 + 3 358 = 42 179 ménages de l'échantillon n'ayant fourni aucune réponse aux questions-échantillon ont été

retranchés de l'échantillon (c'est-à-dire qu'ils ont été convertis de ménages 2B en ménages 2A pour devenir des ménages non inclus dans l'échantillon) et seules les réponses aux questions de base ont été imputées. Ce processus de conversion des ménages inclus dans l'échantillon en ménages non inclus dans l'échantillon est appelé conversion de document 2A/2B. Un petit nombre (52) de ménages 2A et 4 ont été convertis en ménages 2B en vue de satisfaire à certaines contraintes de pondération. Il est possible que les ménages manquants ou refusant de répondre, ainsi que les ménages n'ayant pas répondu aux questions-échantillon, aient des caractéristiques différentes des autres ménages (ils pourraient être, par exemple, de plus petite taille). Ainsi, la conversion de ménages 2B en ménages 2A pourrait biaiser l'échantillon. Par ailleurs, si le système d'imputation avait tendance à imputer plus souvent certaines caractéristiques aux ménages manquants ou refusant de répondre qu'à d'autres types de ménages, des divergences entre les estimations-échantillon et les chiffres de population auraient été produites, car seuls des ménages non inclus dans l'échantillon seraient touchés.

La variation géographique du biais a aussi été étudiée. Le tableau 5.3 montre les valeurs Z correspondant aux 32 caractéristiques pour l'ensemble du Canada, la région de l'est, le Québec, l'Ontario et la région de l'ouest. Bien qu'il soit difficile de tirer des conclusions définitives de ce tableau, étant donné que la taille des échantillons varie énormément, certaines tendances sont évidentes. Le Québec se distingue des autres régions en ce qui concerne les hommes. En effet, ceux-ci sont surreprésentés dans l'échantillon du Québec, alors qu'ils sont sous-représentés dans le reste du pays. Cependant, l'Ontario se démarque également des autres régions pour ce qui est de la population totale. Le nombre de célibataires et de personnes vivant en union libre est aussi surreprésenté au Québec, tandis qu'il est sous-représenté dans le reste du Canada. Mentionnons que ce résultat reflète peut-être davantage une différence dans la façon dont l'état matrimonial est enregistré au Québec. Il est à noter qu'il y a concordance quant aux groupes d'âge d'une région à l'autre, pour ce qui est de l'orientation du biais, exception faite de quelques différences mineures pour les personnes d'âge moyen. Les personnes âgées de plus de 15 ans sont représentées à peu près de la même façon dans toutes les régions, les hommes étant sous-représentés et les femmes surreprésentées (sauf en Ontario, où les femmes sont très légèrement sous-représentées). Enfin, le biais suit également une tendance assez constante en ce qui touche la taille des ménages. Dans l'ensemble, malgré les quelques différences observées entre les régions, on observe de nombreuses similitudes.

Enfin, le tableau 5.4 présente brièvement quelques résultats comparatifs entre l'Examen du biais d'échantillonnage de 1991 et celui de 1996. La comparaison des résultats au niveau des DR révèle que l'orientation du biais n'a changé que dans le cas des hommes et des célibataires. En 1991, ces catégories de personnes étaient surreprésentées dans l'échantillon, alors qu'elles étaient sous-représentées en 1996. Si l'on compare le nombre de DR affichant des valeurs Z supérieures à 0, on constate une différence tout à fait saisissante entre les deux échantillons en ce qui touche les hommes. En 1991, les hommes étaient surreprésentés dans 63 % des DR, tandis que cette proportion était de 47 % en 1996. Le nombre de DR dans lesquelles les célibataires étaient sous-représentés constitue l'autre changement important : en 1991, les célibataires étaient surreprésentés dans 56 % des DR, comparativement à 47 % en 1996. Cependant, au chapitre des groupes d'âge, les deux échantillons présentaient des similitudes remarquables; seules les personnes âgées de 35 à 44 ans étaient surreprésentées dans un plus grand pourcentage de DR en 1991 par rapport à 1996.

En résumé, il existe des biais dans l'échantillon 2B et ces biais sont semblables à ceux qui ont été observés dans l'échantillon du recensement de 1991. Nous ne pouvons toutefois pas déterminer à quoi ces biais sont attribuables ni ce qu'il conviendrait de faire pour les éliminer. Cependant, la méthode de pondération choisie (la PEGMC) contribuera dans une large mesure à corriger les biais pour les caractéristiques choisies comme contraintes. Reste à savoir quels sont les biais que l'on observera pour les caractéristiques n'ayant pas été retenues comme contraintes.

Tableau 5.1 Statistiques sommaires au niveau du Canada

Caractéristique	Chiffres	Estimation[1]	Différence[2]	Écart[3]	E.T.[4]	Valeur Z[5]
Hommes	13 717 654	13 694 786	(22 868)	-0,17	5 752	-3,98
Femmes	14 176 680	14 222 665	45 985	0,32	5 552	8,28
Population totale	**27 894 334**	**27 917 451**	**23 117**	**0,08**	**8 227**	**2,81**
Âge 0 à 4	1 858 332	1 874 111	15 779	0,85	3 073	5,14
Âge 5 à 9	1 932 023	1 950 728	18 705	0,97	3 120	6,00
Âge 10 à 14	1 939 776	1 957 694	17 918	0,92	3 125	5,73
Âge 15 à 19	1 903 023	1 907 732	4 709	0,25	3 074	1,53
Âge 20 à 24	1 840 654	1 816 301	(24 353)	-1,32	3 013	-8,08
Âge 25 à 29	1 971 123	1 953 292	(17 831)	-0,90	3 053	-5,84
Âge 30 à 34	2 405 559	2 401 580	(3 979)	-0,17	3 317	-1,20
Âge 35 à 39	2 486 060	2 482 136	(3 924)	-0,16	3 339	-1,18
Âge 40 à 44	2 268 423	2 273 674	5 251	0,23	3 177	1,65
Âge 45 à 49	2 050 229	2 059 233	9 004	0,44	3 040	2,96
Âge 50 à 54	1 581 484	1 589 751	8 267	0,52	2 707	3,05
Âge 55 à 59	1 271 221	1 269 086	(2 135)	-0,17	2 448	-0,87
Âge 60 à 64	1 157 926	1 160 459	2 533	0,22	2 338	1,08
Âge 65 à 74	1 991 721	1 996 303	4 582	0,23	3 068	1,49
Âge 75 et plus	1 236 780	1 225 372	-11 408	-0,92	2 332	-4,89
Personnes célibataires	12 779 218	12 741 878	(37 340)	-0,29	7 320	-5,10
Personnes mariées	11 537 475	11 628 813	91 338	0,79	6 076	15,03
Veufs et veuves	1 303 304	1 291 501	(11 803)	-0,91	2 130	-5,54
Personnes divorcées	1 605 136	1 591 530	-13 606	-0,85	2 612	-5,21
Personnes séparées	669 201	663 729	-5 472	-0,82	1 675	-3,27
Union libre = oui	1 770 338	1 768 774	-1 564	-0,09	3 568	-0,44
Hommes ≥ 15	10 781 073	10 732 804	-48 269	-0,45	4 449	-10,85
Femmes ≥ 15	11 383 130	11 402 113	18 983	0,17	4 006	4,74
Ménages d'une personne	2 584 348	2 558 041	-26 307	-1,02	2 524	-10,42
Ménages de deux personnes	3 385 597	3 397 657	12 060	0,36	3 011	4,00
Ménages de trois personnes	1 804 304	1 809 076	4 772	0,26	2 435	1,96
Ménages de quatre personnes	1 813 493	1 825 159	11 666	0,64	2 378	4,91
Ménages de cinq personnes	737 751	740 921	3 170	0,43	1 640	1,93
Ménages de six personnes ou plus	334 207	327 786	-6 421	-1,92	1 124	-5,71

[1] Fondée sur les coefficients de pondération simples
[2] Différence : estimation-chiffre
[3] Écart : écart (100*[différence-chiffre]/chiffre)
[4] E.T. : erreur-type de l'estimation du coefficient de pondération simple
[5] Valeur Z : (estimation-biais-chiffre)/E.T.

Tableau 5.2 Statistiques sommaires des valeurs Z au niveau des divisions de recensement (DR)

Caractéristique	Moyenne	Type[1]	T[2]	% DR (Z > 2)	% DR (Z < -2)	% DR (Z > 0)
Hommes	-0,07	1,08	-1,10	2,1	3,6	47,3
Femmes	0,55	1,08	8,50	8,2	0,4	69,7
Population totale	0,32	1,06	4,98	6,0	0,7	59,8
Âge 0 à 4	0,33	1,06	5,26	6,4	2,5	60,1
Âge 5 à 9	0,33	1,04	5,29	4,3	0,7	61,2
Âge 10 à 14	0,29	0,99	4,87	3,9	0,7	60,1
Âge 15 à 19	0,10	0,98	1,73	3,6	0,7	54,8
Âge 20 à 24	-0,38	1,04	-6,20	1,1	7,1	37,4
Âge 25 à 29	-0,13	1,07	-2,03	1,8	2,8	48,4
Âge 30 à 34	0,15	1,02	2,38	3,2	2,8	60,1
Âge 35 à 39	0,03	0,98	0,60	2,1	1,8	53,7
Âge 40 à 44	0,13	1,01	2,10	3,6	0,7	53,4
Âge 45 à 49	0,22	1,10	3,29	5,3	2,1	59,1
Âge 50 à 54	0,13	1,03	2,12	4,3	2,1	55,5
Âge 55 à 59	-0,05	0,99	-0,78	1,8	2,1	49,8
Âge 60 à 64	0,05	1,10	0,82	2,5	4,6	53,4
Âge 65 à 74	-0,02	1,03	-0,40	1,8	2,1	50,5
Âge 75 et plus	-0,43	1,01	-7,11	0,4	7,8	31,0
Personnes célibataires	-0,13	1,10	-1,91	1,1	3,9	46,6
Personnes mariées	0,87	1,03	14,15	12,5	0,4	81,9
Veufs et veuves	-0,41	0,96	-7,06	0,4	4,3	31,7
Personnes divorcées	-0,23	1,02	-3,78	0,7	6,0	42,7
Personnes séparées	-0,21	0,94	-3,68	1,4	2,1	39,9
Union libre = oui	-0,02	0,91	-0,33	0,7	1,4	52,0
Hommes ≥ 15	-0,39	1,19	-5,50	1,8	6,8	36,7
Femmes ≥ 15	0,34	1,08	5,22	6,4	1,1	62,3
Ménages d'une personne	-0,69	1,02	-11,37	0,0	9,6	26,3
Ménages de deux personnes	0,20	1,04	3,18	5,3	2,1	55,9
Ménages de trois personnes	0,15	0,98	2,54	3,2	0,7	54,8
Ménages de quatre personnes	0,30	1,12	4,44	3,2	2,5	62,3
Ménages de cinq personnes	0,16	0,96	2,74	1,1	1,4	56,9
Ménages de six personnes ou plus	-0,26	1,03	-4,18	1,4	5,3	40,9

[1] Type : déviation type des valeurs Z
[2] T : statistique étudiante pour tester que moyenne = 0

Tableau 5.3 Valeurs Z aux niveaux national et régional

Caractéristique	Canada	Est	Québec	Ontario	Ouest
Hommes	-3,98	-0,37	2,48	-5,69	-2,91
Femmes	8,28	3,87	7,57	1,82	4,29
Population totale	2,81	2,34	6,94	-2,75	0,89
Âge 0 à 4	5,14	0,33	4,37	2,10	2,91
Âge 5 à 9	6,00	3,75	4,67	1,42	3,26
Âge 10 à 14	5,73	1,61	2,88	2,88	3,81
Âge 15 à 19	1,53	1,03	2,03	0,24	0,15
Âge 20 à 24	-8,08	-2,46	-2,35	-4,78	-5,97
Âge 25 à 29	-5,84	-0,81	-0,26	-6,31	-2,84
Âge 30 à 34	-1,20	2,15	1,78	-3,51	-0,92
Âge 35 à 39	-1,18	0,36	-0,52	-0,98	-0,77
Âge 40 à 44	1,65	1,30	2,10	-0,79	1,30
Âge 45 à 49	2,96	1,47	2,15	1,22	1,31
Âge 50 à 54	3,05	-0,52	2,29	2,19	1,29
Âge 55 à 59	-0,87	1,41	-0,33	-1,96	0,20
Âge 60 à 64	1,08	0,11	3,02	-1,00	0,26
Âge 65 à 74	1,49	-0,68	-1,02	2,93	0,70
Âge 75 et plus	-4,89	-3,54	-2,73	-1,46	-2,95
Personnes célibataires	-5,10	-0,58	0,87	-6,55	-2,69
Personnes mariées	-15,03	6,10	9,16	7,12	8,03
Veufs et veuves	-5,54	-2,66	-2,01	-2,98	-3,61
Personnes divorcées	-5,21	-1,98	-0,98	-3,60	-3,85
Personnes séparées	-3,27	-1,51	-0,30	-2,28	-2,26
Union libre = oui	-0,44	-0,54	1,09	-1,23	-0,74
Hommes ≥ 15	-10,85	-2,31	-1,27	-9,23	-7,19
Femmes ≥ 15	4,74	3,04	6,27	-0,32	1,75
Ménages d'une personne	-10,42	-5,02	-8,38	-3,20	-5,01
Ménages de deux personnes	4,00	1,70	0,68	3,32	2,17
Ménages de trois personnes	1,96	0,15	2,79	0,55	0,26
Ménages de quatre personnes	4,91	2,76	4,32	1,17	2,29
Ménages de cinq personnes	1,93	-0,02	2,43	-0,21	1,71
Ménages de six personnes ou plus	-5,71	-0,74	-0,80	-5,18	-3,12

Tableau 5.4 Comparaison entre 1991 et 1996

Caractéristique	1996			1991		
	T-Test	Moyenne	% DR > 0	T-Test	Moyenne	% DR > 0
Hommes	-1,1	-0,1	47	5,1	0,3	63
Femmes	8,5	0,6	70	12,0	0,8	78
Population totale	5,0	0,3	60	11,2	0,7	76
Âge 0 à 4	5,3	0,3	60	5,2	0,3	63
Âge 5 à 9	5,3	0,3	61	7,6	0,5	68
Âge 10 à 14	4,9	0,3	60	8,2	0,5	64
Âge 15 à 19	1,7	0,1	55	3,4	0,2	59
Âge 20 à 24	-6,2	-0,4	37	-3,6	-0,2	43
Âge 25 à 29	-2,0	-0,1	48	-2,5	-0,2	46
Âge 30 à 34	2,4	0,2	60	3,2	0,2	60
Âge 35 à 39	0,6	0,0	54	3,0	0,2	62
Âge 40 à 44	2,1	0,1	53	4,3	0,3	61
Âge 45 à 49	3,3	0,2	59	4,3	0,3	62
Âge 50 à 54	2,1	0,1	56	2,3	0,1	58
Âge 55 à 59	-0,8	-0,1	50	-0,1	0,0	51
Âge 60 à 64	0,8	0,1	53	1,0	0,1	51
Âge 65 à 74	-0,4	0,0	51	-0,2	-0,0	49
Âge 75 et plus	-7,1	-0,4	31	-6,6	-0,4	35
Personnes célibataires	-1,9	-0,1	47	3,2	0,2	56
Personnes mariées	14,2	0,9	82	17,4	1,1	85
Veufs et veuves	-7,1	-0,4	32	-6,1	-0,4	37
Personnes divorcées	-3,8	-0,2	43	-5,2	-0,3	38
Personnes séparées	-3,7	-0,2	40	-5,2	-0,3	39
Ménages d'une personne	-11,4	-0,7	26	-17,1	-1,0	15
Ménages de deux personnes	3,2	0,2	56	3,1	0,2	58
Ménages de trois personnes	2,5	0,2	55	3,6	0,2	59
Ménages de quatre personnes	4,4	0,3	62	7,6	0,5	69
Ménages de cinq personnes	2,7	0,2	57	4,6	0,3	58
Ménages de six personnes ou plus	-4,2	-0,3	41	-1,4	-0,1	43

VI. Évaluation des méthodes de pondération

A. Formation des régions de pondération (RP)

La première étape du processus de pondération a été la formation des RP. Une RP est la plus petite unité géographique pour laquelle une concordance des caractéristiques de la population entre certains résultats-échantillon et chiffres de population peut être assurée. La formation de RP plus grandes permet une meilleure cohérence entre les chiffres de population et les estimations pour un plus grand nombre de caractéristiques démographiques, mais au détriment de la cohérence entre les chiffres de population et les estimations pour les petites régions. En conséquence, en ce qui concerne la taille des RP (nombre de ménages), nous devons faire un compromis entre la nécessité de faire concorder les estimations-échantillon et les chiffres de population des petites régions, d'une part, et la nécessité de faire en sorte que cette concordance s'applique à un vaste éventail de caractéristiques. Pour le recensement de 1996, la formation des RP, qui consistait à regrouper des secteurs de dénombrement (SD), devait respecter les conditions suivantes :

a) Une RP doit respecter les limites des divisions de recensement (DR).

b) Une RP devrait contenir entre 800 et 3 000 ménages.

c) Une RP devrait, dans la mesure du possible, respecter les limites de certaines régions de niveau supérieur, la priorité étant donnée, par ordre d'importance, aux subdivisions de recensement (SDR), aux secteurs de recensement (SR) et, enfin, aux circonscriptions électorales fédérales (CÉF).

d) Les RP devraient être formées de SD contigus (c.-à-d. être continues) et être le plus compacte possible.

Étant donné que la Procédure d'estimation généralisée des moindres carrés (PEGMC) est effectuée de façon indépendante dans chaque RP, on obtiendra une concordance entre les estimations-échantillon et les chiffres de population uniquement pour les régions géographiques qui renferment exclusivement des RP complètes. La concordance n'est pas assurée pour les régions géographiques faisant intégralement partie d'une portion de RP ou qui contiennent des portions de différentes RP.

Le tableau suivant montre la répartition des RP selon leur taille (nombre de ménages).

Tableau 6.1 Répartition des régions de pondération de 1996 selon le nombre de logements et la contiguïté

Logements	RP
0 à 999	4
1 000 à 1 499	1 686
1 500 à 1 999	2 213
2 000 à 2 499	1 417
2 500 à 2 999	560
3 000+	61
Total	**5 941**

Le tableau 6.1 montre que 5 876 RP (98,9 %) se situaient à l'intérieur de l'intervalle souhaité (1 000 à 3 000 ménages) en 1996. Par ailleurs, 5 888 RP (99,1 %) étaient contiguës. La plupart des autres RP n'étaient pas contiguës en raison de la non-contiguïté d'un SD ou d'une SDR.

Le tableau suivant montre le nombre de régions géographiques dont les limites respectent celles des RP.

Tableau 6.2 Nombre de SDR, SR et CÉF qui respectent les limites de RP

Description	Limites géographiques de 1996		
	SDR	SR	CÉF
Régions géographiques renfermant seulement une RP complète ou une partie d'une seule RP	5 373 (89,8 %)	4 396 (73,5 %)	0 (0 %)
Régions géographiques renfermant plus d'une RP complète	329 (5,5 %)	313 (5,2 %)	46 (15,6 %)
Régions géographiques chevauchant au moins une limite de RP	282 (4,7 %)	1 272 (21,3 %)	249 (84,4 %)
Total	**5 984**	**5 980**	**295**

Le tableau indique que 282 SDR (4,7 %), 1 272 SR (21,3 %) et 249 CÉF (84,4 %) ne respectent pas les limites des RP.

Pour obtenir de plus amples renseignements au sujet des régions de pondération et de leurs limites, consulter Kruszynski, 1999.

B. Évaluation de la Procédure d'estimation généralisée des moindres carrés

Le système de pondération de 1991 a été utilisé en 1996 et le logiciel n'a été que très légèrement modifié. Bien que cela suppose que l'on a apporté peu de changement à l'ensemble du système, il n'en reste pas moins que les paramètres (les contraintes définies, par exemple) pouvaient être modifiés. L'objectif était de conserver une proportion plus élevée de contraintes qu'on ne l'avait fait en 1991. En outre, en raison de restrictions budgétaires, certaines des caractéristiques 2A utilisées en 1991 ont été éliminées.

Durant le premier exercice de production, les contraintes associées à la taille des ménages ont souvent été abandonnées au niveau des RP, ce qui s'est traduit par une différence entre les chiffres de population et les estimations pour ces contraintes. On a observé un écart important de 2,6 % pour la caractéristique **« Union libre = oui »** (qui n'avait pas été utilisée comme contrainte lors du premier exercice). Ces observations donnent à penser que le système de pondération n'a pas réussi à éliminer complètement les biais dans l'échantillon du recensement et que les coefficients de régression pourraient alors donner lieu à des estimations biaisées en ce qui concerne les caractéristiques 2B. C'est pourquoi il a été décidé de mettre le système de pondération à l'essai une deuxième fois de la façon décrite dans le chapitre III, section D du présent rapport.

Pour réduire l'écart associé à la caractéristique **« Union libre = oui »**, celle-ci a été ajoutée comme contrainte lors du deuxième exercice. Par ailleurs, en vue de réduire les écarts observés quant à la taille des ménages, on a ajusté les coefficients de pondération initiaux de manière à obtenir une concordance avec les chiffres de population enregistrés pour les diverses tailles de ménages au niveau des RP. Lors du premier exercice, les coefficients de pondération initiaux pour tous les ménages d'une RP donnée correspondaient simplement au nombre de ménages de la RP divisé par le nombre de ménages échantillonnés dans la RP : ces coefficients de pondération sont appelés **coefficients de pondération initiaux simples**. Lors du second exercice, les coefficients de pondération initiaux simples ont été ajustés pour concorder avec le nombre estimatif de ménages – pour chacune des six tailles de ménage (1, 2, 3, 4, 5, 6+) – au niveau des RP. Ainsi, au départ, avant de mettre à l'essai le système de pondération par régression, une concordance a été établie au niveau des RP entre les estimations-échantillon et les chiffres de population pour les six tailles de ménage. Ces coefficients de pondération ajustés sont appelés **coefficients de pondération initiaux stratifiés a posteriori**. Le système de pondération par régression a ensuite été appliqué à ces coefficients de pondération. Pour obtenir de plus amples renseignements, consulter Bankier, Houle et Luc, 1997.

L'évaluation du système de pondération s'est déroulée en quatre étapes. Premièrement, la répartition des coefficients de pondération définitifs a été examinée. Deuxièmement, les écarts entre les chiffres de population et les estimations par régression finales ont été affichés et comparés aux chiffres correspondants de 1991. Troisièmement, les contraintes qui avaient été retranchées lors du premier exercice ont été décrites brièvement, puis tous les détails associés aux contraintes retranchées lors du second exercice ont été affichés. Enfin, le rendement du système de pondération a été évalué aux divers niveaux géographiques.

1. Répartition des coefficients de pondération

Le tableau 6.3 montre la répartition des coefficients de pondération pour les deux exercices de production réalisés en 1996 et la répartition correspondante pour 1991. Il est à noter que, en vue de conserver un plus grand nombre de contraintes qu'en 1991, il a été décidé de permettre l'utilisation de coefficients de pondération inclus dans l'intervalle de 0 à 1, de sorte que les intervalles allaient de 0 à 25 en 1996 alors qu'ils allaient de 1 à 25 en 1991.

Tableau 6.3 Répartition des coefficients de pondération de ménages

Intervalles des coefficients de pondération	1991		1996 – Deuxième		1996 – Première	
	Nombre	%	Nombre	%	Nombre	%
< 1,00	0	-	13 833	0,7	13 611	0,6
1,00 à 1,99	58 934	3,1	70 195	3,4	68 116	3,3
2,00 à 2,99	170 352	8,9	204 749	9,8	198 052	9,5
3,00 à 3,99	338 334	17,7	383 921	18,4	378 736	18,1
4,00 à 4,99	432 603	22,6	461 281	22,1	465 983	22,3
5,00 à 5,99	369 274	19,3	379 204	18,1	388 137	18,6
6,00 à 6,99	242 000	12,7	245 652	11,8	250,521	12,0
7,00 à 7,99	137 472	7,2	143 887	6,9	144 662	6,9
8,00 à 8,99	74 270	3,9	80 913	3,9	80 486	3,9
9,00 à 9,99	39 671	2,1	44 793	2,1	43 947	2,1
10,00 à 14,99	45 808	2,4	56 569	2,7	53 348	2,6
15,00 à 19,99	3 182	0,2	4 660	0,2	4 099	0,2
20,00 à 24,99	357	0,0	525	0,0	484	0,0
≥ 25,00	0	-	0	-	0	-
Total	**1 912 257**	**100,0**	**2 090 182**	**100,0**	**2 090 182**	**100,0**

Le tableau révèle que seule une petite proportion de ménages, soit moins de 0,7 %, avaient un coefficient de pondération inférieur à 1. Dans l'ensemble, on observe une plus forte proportion de coefficients de pondération élevés ou faibles en 1996 qu'en 1991. Mentionnons qu'en 1996 tous les intervalles inférieurs à 3 comptaient pour un pourcentage plus important du total que le groupe correspondant en 1991. En outre, tous les intervalles de 9 ou plus équivalaient également à un pourcentage plus élevé que le groupe correspondant en 1991. Cependant, pour ce qui est des intervalles intermédiaires, entre 4 et 9, l'échantillon de 1991 renfermait un pourcentage plus élevé de ménages que celui de 1996. Si on examine le pourcentage de ménages correspondant aux intervalles situés entre 3 et 7 en 1986 (Bankier et coll., 1992) et en 1991 par rapport à 1996, on constate une diminution marquée. En 1986, alors que l'on utilisait la méthode itérative du quotient, on trouvait 93,9 % des ménages dans ces intervalles; en 1991, ce pourcentage était tombé à 72,3 %, puis il est descendu encore en 1996 pour se chiffrer à 71 % (premier exercice) et à 70,3 % (deuxième exercice).

2. Écart entre les chiffres de population et les estimations-échantillon

Un des buts de la méthode de pondération consiste à réduire les écarts entre les chiffres de population et les estimations-échantillon correspondantes pour les contraintes. Ces écarts sont le résultat de la variabilité et du biais d'échantillonnage (voir le chapitre V). Certains écarts peuvent toutefois subsister même après l'application de la méthode de pondération. Les écarts correspondent à la différence entre les estimations-échantillon et les chiffres de population, laquelle est exprimée en tant que pourcentage des chiffres de population, c'est-à-dire :

$$\textbf{écart} = \frac{\text{estimation-échantillon} - \text{chiffre de population}}{\text{chiffre de population}} \times 100 \qquad (2)$$

Le numérateur de cette expression (estimation-échantillon – chiffre de population) est souvent appelé la « **différence** ». On notera que les chiffres de population correspondent aux SD échantillonnés uniquement.

Le tableau 6.4 montre les **différences** et les **écarts** observés à l'échelle du Canada en 1996 pour 32 caractéristiques démographiques de base. On présente les **différences** pour trois estimations-échantillon distinctes, à savoir les coefficients de pondération initiaux, les coefficients de pondération établis lors du premier exercice et les coefficients de pondération établis lors du deuxième exercice. Soulignons que les écarts ont été arrondis à deux décimales. Les estimations-échantillon et les chiffres de population sont fondés sur les logements privés occupés dans les SD échantillonnés.

Le tableau montre clairement que, lors du premier exercice, le système de pondération a permis de réduire les différences pour la plupart des caractéristiques. Ainsi, les coefficients de pondération initiaux ont produit un écart de 91 338 entre les estimations-échantillon et les chiffres de population pour les personnes mariées, tandis que lors du premier exercice de pondération, on a réduit cet écart à 10. Les écarts associés aux groupes d'âge plus jeunes ont aussi diminué considérablement. Le système de pondération n'a toutefois pas permis de réduire l'écart pour toutes les caractéristiques. En fait, pour ce qui est de la caractéristique « union libre », les coefficients de pondération initiaux ont donné une différence de -1 404, que le système de pondération a fait passer à 46 646. Des différences similaires ont également été observées au chapitre de la taille des ménages, en particulier pour ce qui est des ménages de cinq et six personnes. Mentionnons que, dans le cas des ménages de cinqpersonnes, les coefficients de pondération initiaux ont donné une différence de 3 170, que le système de pondération a fait passer à 57 775. Ces résultats constituent la principale raison pour laquelle l'exercice a été effectué une deuxième fois.

Comme nous l'avons déjà mentionné, lors du deuxième exercice de pondération, la caractéristique « union libre » a été utilisée comme contrainte. En outre, les coefficients de pondération initiaux ont été ajustés de manière à obtenir une concordance avec les chiffres de population associés à la taille des ménages au niveau des RP. Le tableau 6.4 fait également état des différences observées lors du deuxième exercice. Les principales difficultés rencontrées lors du premier exercice ont été résolues lors du deuxième. Ainsi, l'écart associé à la caractéristique « union libre » est tombé de 46 646 à 2 415 et la différence enregistrée pour les ménages de cinq personnes a chuté de 57 775 à 27 879. Cependant, des résultats un peu plus élevés ont été enregistrés pour certaines caractéristiques lors du deuxième exercice. Les personnes âgées de 75 ans et plus, par exemple, affichaient une différence de -2 377 lors du premier exercice, mais cette différence est passée à -9 207 lors du deuxième.

Tableau 6.4 Écarts estimation/population à l'échelle du Canada en 1996

Contrainte	Coefficients de pondération initiaux	Premier exercice		Deuxième exercice	
	Différence	Différence	Écart	Différence	Écart
Hommes	-22 868	-	0,00	15	0,00
Hommes de plus de 14 ans	-48 269	213	0,00	-276	0,00
Personnes de plus de 14 ans	-29 285	3	0,00	3	0,00
Total des ménages	1 060	-	0,00	-	0,00
Population totale	**23 117**	-	**0,00**	-	**0,00**
Âge 0 à 4	15 779	34	0,00	-208	-0,01
Âge 5 à 9	18 705	168	0,01	-258	-0,01
Âge 10 à 14	17 918	-205	-0,01	462	0,02
Âge 15 à 19	4 709	1 890	0,10	1 853	0,10
Âge 20 à 24	-24 353	1 501	0,08	803	0,04
Âge 25 à 29	-17 831	-49	-0,00	105	0,01
Âge 30 à 34	-3 979	248	0,01	361	0,02
Âge 35 à 39	-3 924	493	0,02	320	0,01
Âge 40 à 44	5 251	232	0,01	366	0,02
Âge 45 à 49	9 004	1 743	0,09	971	0,05
Âge 50 à 54	8 267	959	0,06	993	0,06
Âge 55 à 59	-2 135	-201	-0,02	254	0,02
Âge 60 à 64	2 533	-3 380	-0,29	3 847	0,33
Âge 65 à 74 #	4 582	-1 056	-0,05	-662	-0,03
Âge 75 et plus	-11 408	-2 377	-0,19	-9 207	-0,74
Personnes célibataires	-37 340	239	0,00	115	0,00
Personnes mariées	91 338	10	0,00	73	0,00
Veufs et veuves	-11 803	-1 338	-0,10	-1 387	-0,11
Personnes divorcées	-13 606	951	0,06	1 209	0,08
Personnes séparées	-5 472	137	0,02	-10	-0,00
Union libre = oui	-1 404	46 646	2,63	2 415	0,14
Ménages d'une personne #	-	-	-	-4 750	-0,18
Ménages de deux personnes	12 060	-344	-0,01	-3 331	-0,05
Ménages de trois personnes	4 772	3 317	0,06	2 614	0,05
Ménages de quatre personnes	11 666	4 822	0,07	6 776	0,09
Ménages de cinq personnes	3 170	57 775	1,57	27 879	0,76
Ménages de six personnes ou plus #	-	-65 570	-1,38	-29 187	-0,52

= Contrainte ajoutée pour le deuxième exercice.

3. Retranchement des contraintes

Lors du premier exercice, 29 contraintes (voir l'annexe B) au total ont été utilisées au niveau des RP, y compris les groupes d'âge quinquennaux, l'état matrimonial, le sexe et la taille du ménage. La contrainte « État matrimonial – Séparé » n'a pas été utilisée, car il s'agit d'une contrainte linéairement dépendante aux autres contraintes associées à l'état matrimonial. De la même façon, les contraintes « Âge 60 à 64 » et « Ménage de six personnes ou plus » n'ont pas été utilisées parce qu'elles sont linéairement dépendantes à d'autres contraintes. Comme nous l'avons mentionné précédemment, en vue de conserver un plus grand nombre de contraintes, nous avons utilisé en 1996 des coefficients de pondération se situant dans un intervalle de 0 à 25, au lieu d'un intervalle de 1 à 25 comme en 1991.

Le tableau 6.5 présenté ci-après montre la fréquence à laquelle les contraintes ont été retranchées lors du deuxième exercice au niveau des RP et les raisons s'y rattachant. Ainsi, comme une contrainte **linéairement dépendante (LD)** est redondante (ménages de six personnes ou plus, par exemple, puisque l'ensemble des ménages est déjà une

contrainte; par conséquent, l'un des sous-groupes doit être redondant), le fait de la retrancher n'aura aucune incidence sur les estimations finales. On appelle contraintes **faibles**, les contraintes qui s'appliquent à moins de 20 ménages, tandis que les contraintes **quasi linéairement dépendantes (QLD)** et les **observations aberrantes** correspondent à des coefficients de pondération supérieurs à 25 ou inférieurs à 0.

Le tableau montre le nombre de fois où les contraintes ont été retranchées. Ainsi, pour un total de 5 941 RP, la contrainte « Âge 0 à 4 » a été retranchée 3 154 fois, ce qui laisse supposer qu'il existe un risque d'écart entre le chiffre de population et l'estimation-échantillon obtenu à l'aide de coefficients de pondération par régression. L'écart associé à cette caractéristique est très faible (-208 personnes de 0 à 4 ans, selon le tableau 6.4), car il s'agit d'une contrainte **linéairement dépendante**. Cependant, en ce qui concerne les ménages d'une personne, la différence augmente, car cette contrainte a été retranchée parce qu'il s'agit d'une contrainte **quasi linéairement dépendante**. Il convient de noter que, dans le cas des ménages d'une personne, la contrainte a été retranchée 4 600 fois (dans 4 583 cas, parce qu'il s'agissait d'une contrainte QLD) et que, par conséquent, l'écart entre le chiffre de population et l'estimation obtenue au moyen de coefficients de pondération par régression est de -4 740. Il existe un lien évident entre le nombre de fois où une contrainte est retranchée et la différence qui en résulte, pourvu que la contrainte ne soit pas retranchée en raison de sa dépendance linéaire.

Il existe un groupe de caractéristiques dont les résultats sont différents. Soulignons que les caractéristiques « Âge 0 à 4 », « Âge 5 à 9 » et « Âge 10 à 14 » ont toutes été retranchées à une fréquence plutôt élevée. Cependant, les contraintes « Personnes de plus de 15 ans » et « Population totale » sont presque toujours retenues. Ainsi, l'important groupe d'âge de 0 à 14 ans constitue une contrainte redondante, ce qui explique pourquoi les trois groupes d'âge sont retranchés avec une telle régularité en raison de leur dépendance linéaire.

Tableau 6.5 Fréquence à laquelle les contraintes au niveau des RP ont été retranchées en 1996

Caractéristique	Faible	LD	QLD	Observation aberrante	Total
Hommes	0	0	0	1	1
Femmes*	-	-	-	-	-
Population totale	**0**	**0**	**0**	**0**	**0**
Âge 0 à 4	6	3 071	20	57	3 154
Âge 5 à 9	30	709	77	135	951
Âge 10 à 14	35	2 110	33	61	2 239
Âge 15 à 19	6	514	27	96	643
Âge 20 à 24	1	216	133	119	469
Âge 25 à 29	1	347	108	82	538
Âge 30 à 34	1	29	23	42	95
Âge 35 à 39	1	0	6	31	38
Âge 40 à 44	1	3	13	45	62
Âge 45 à 49	1	4	9	50	64
Âge 50 à 54	2	157	67	83	309
Âge 55 à 59	2	636	213	147	998
Âge 60 à 64	3	1 122	973	128	2 226
Âge 65 à 74	4	3	214	81	302
Âge 75 et plus	36	2 864	100	60	3 060
Personnes célibataires	0	0	1	3	4
Personnes mariées	0	0	0	4	4
Veufs et veuves	2	0	174	345	521
Personnes divorcées	1	1	213	252	467
Personnes séparées*	-	-	-	-	-
Union libre = oui	23	0	1	272	296
Ménages d'une personne	1	12	4 583	4	4 600
Ménages de deux personnes	0	0	1 154	12	1 166
Ménages de trois personnes	2	22	189	47	260
Ménages de quatre personnes	23	145	52	37	257
Ménages de cinq personnes	193	997	865	92	2 147
Ménages de six personnes ou plus *	-	-	-	-	-
Hommes $\geq$ 15 ans	0	1	136	3	140
Personnes $\geq$ 15 ans	0	0	1	0	1
Total des ménages	0	0	0	0	0

* Indique que la caractéristique n'a pas été utilisée comme contrainte parce qu'il s'agissait d'une contrainte redondante.

4. Évaluation des divers niveaux géographiques

Une étude a été faite pour comparer les différences absolues entre les estimations-échantillon et les chiffres de population pour 31 caractéristiques observées en 1996 (29 de celles-ci étaient des contraintes) et en 1986, à différents niveaux géographiques. Les 31 caractéristiques incluses dans cette étude des différences absolues sont présentées à l'annexe B. Les résultats de l'étude sont résumés au tableau 6.6 ci-dessous. Le tableau indique le pourcentage de caractéristiques ayant une « valeur R » se situant à l'intérieur d'un certain intervalle pour les six niveaux géographiques qui figurent dans le tableau. Une valeur R est un quotient des différences de 1996 et celles de 1986 calculé selon l'expression suivante :

$$R = 100 * \frac{\sum_{1}^{N_{96}} \left|\hat{X}^{96} - X^{96}\right| / \sum_{1}^{N_{86}} \left|X^{86}\right.}{\sum_{1}^{N_{96}} \left|\hat{X}^{96} - X^{96}\right| / \sum_{1}^{N_{86}} \left|X^{86}\right.} \qquad (3)$$

où X^{96} et X^{86} sont respectivement les chiffres de population de 1996 et 1986 pour une caractéristique. L'estimation-échantillon de 1996 d'après les coefficients de pondération PEGMC est $\hat{X}^{96}$, tandis que l'estimation-échantillon de 1986 fondée sur les coefficients de pondération MIQ est $\hat{X}^{86}$. Des valeurs R ont été calculées pour chacun des six niveaux géographiques (SD, RP, SDR, DR, province et Canada). La somme des valeurs absolues des différences population/estimation a été calculée, N_{96} étant égal au nombre d'unités du niveau géographique visé en 1996 et N_{86} étant égal au nombre d'unités du niveau géographique visé en 1991. Une valeur R se situant entre 95 et 105 indique que les systèmes d'estimation de 1996 et de 1991 ont donné à peu près le même résultat. Une valeur R inférieure à 95 indique que le système de 1996 a été plus efficace que celui de 1991 pour la caractéristique visée au niveau géographique en question, tandis qu'une valeur R supérieure à 105 indique qu'il a été moins efficace. Le tableau présente également les résultats de 1991, alors que la comparaison avait été établie pour 62 caractéristiques (dont 49 étaient des contraintes).

Tableau 6.6 Pourcentage des caractéristiques ayant des valeurs R dans certains intervalles

	Valeur R	SD	RP	SDR	DR	Province	Canada
1996 vs 1986	< 95	94	64	94	64	52	49
	95 à 105	0	15	0	15	21	18
	> 105	6	21	6	21	27	33
1991 vs 1986	< 95	87	58	81	47	31	29
	95 à 105	11	11	8	18	14	10
	> 105	2	31	11	35	55	61

La méthode d'estimation utilisée en 1996 a permis de réduire les différences entre les chiffres de population et les estimations à tous les niveaux comparativement à la méthode de 1986. Il est à noter qu'en 1991 la méthode d'estimation a permis de réduire les différences au niveau des petites régions géographiques, du SD jusqu'à la DR, mais qu'aux niveaux provincial et national, la méthode de 1986 a donné de plus petites différences.

VII. Examen de la cohérence entre les estimations-échantillon et les chiffres de population

Pour que la PEGMC soit efficace, il a fallu retrancher certaines contraintes dans chaque RP avant de calculer les coefficients de pondération. Plusieurs caractéristiques importantes ont donc été exclues dans un certain nombre de RP. Cela a eu pour conséquence de réduire le niveau de concordance (cohérence) entre les estimations-échantillon et les chiffres de population pour ces caractéristiques. De plus, de nombreuses régions géographiques d'intérêt ne se composent pas toujours de RP complètes. Dans ces régions, la cohérence pour l'ensemble des caractéristiques dépend donc de la mesure dans laquelle les régions se composent de RP complètes.

L'examen de la cohérence a porté sur les écarts entre les estimations-échantillon et les chiffres de population (exprimés en pourcentage des chiffres de population) pour le même ensemble de base de 31 caractéristiques que celles de l'Examen du biais d'échantillonnage (voir l'annexe B) pour les régions géographiques suivantes :

a) divisions de recensement;

b) subdivisions de recensement;

c) secteurs de recensement;

d) secteurs de dénombrement.

Comme il est indiqué au chapitre VI (sous-section B.2), les écarts entre les estimations-échantillon et les chiffres de population ont été calculés de la façon suivante :

$$\textbf{écart} = \frac{\text{Estimation-échantillon} - \text{chiffre de population}}{\text{chiffre de population}} \times 100$$

A. Divisions de recensement (DR)

Les centiles du tableau 7.1 résument les résultats relatifs au niveau de cohérence, pour l'ensemble des DR échantillonnées au Canada, d'une grande variété de caractéristiques de base dont le chiffre de population[6] est supérieur à 50. En général, les écarts obtenus pour des caractéristiques dont le chiffre de population était de ≤ 50 dans la plupart des régions géographiques se sont révélés relativement élevés (soit positifs, soit négatifs). On a donc décidé de ne pas inclure les régions géographiques où le chiffre de population pour la caractéristique était inférieur ou égal à 50, car il suffisait de quelques-unes de ces régions pour perturber sensiblement les centiles des écarts présentés dans les tableaux du présent chapitre. Cette perturbation se produirait si une grande partie de ces régions présentaient soit des écarts positifs relativement élevés, soit des écarts négatifs relativement élevés.

Au tableau 7.1, pour chaque caractéristique, N % des DR présentaient des écarts inférieurs au N[e] centile tandis que 100 – N % des DR avaient des écarts supérieurs au N[e] centile. L'écart se situait donc entre le 10[e] et le 90[e] centile pour 80 % des DR, et entre le 25[e] et le 75[e] centile pour 50 % des DR, etc. Par exemple, l'écart pour le groupe d'âge

[6] Il s'agit du chiffre de population pour la caractéristique. Par exemple, le niveau de cohérence pour le groupe d'âge 0 à 4 est résumé pour toutes les DR dans lesquelles plus de 50 personnes appartenaient au groupe d'âge 0 à 4. La même définition vaut pour les tableaux 7.1, 7.2, 7.3 et 7.4.

0 à 4 variait entre -0,44 % et 0,05 % pour 80 % des DR. Une répartition symétrique de part et d'autre de 0 signifie que les différences entre les estimations-échantillon et les chiffres de population sont réparties de façon égale. Cependant, une répartition non symétrique signifie que les chiffres de l'échantillon ne sont pas répartis de façon uniforme entre les différences positives et négatives.

Toutes les DR sont formées de RP complètes. Ainsi, parmi les caractéristiques qui se sont traduites en contraintes en 1996, celles qui étaient rarement ou jamais retranchées dans une RP affichaient une cohérence quasi parfaite au niveau des DR[7]. Ces caractéristiques étaient les suivantes : nombre total de personnes, nombre total de femmes, nombre total de célibataires, etc. Le niveau de cohérence observé pour les autres caractéristiques n'était pas parfait, mais demeurait quand même très bon, sauf pour les caractéristiques représentant seulement un faible pourcentage de la population dans la plupart des DR, par exemple, le nombre de personnes séparées ou le nombre de ménages de six personnes ou plus. Une relation générale existe entre les écarts et les chiffres de population pour toutes les caractéristiques, c'est-à-dire que la cohérence s'accroît à mesure qu'augmente le chiffre de population pour la DR.

Le tableau 7.1 montre également les centiles correspondant aux écarts de 1991 pour les DR. Les tableaux 7.2, 7.3 et 7.4 contiennent également les données de 1991 pour les autres niveaux géographiques. Si l'on compare les données de 1991 à celles de 1996, on constate que les écarts enregistrés en 1996 au niveau des DR sont identiques ou significativement inférieurs à ceux de 1991 pour toutes les caractéristiques figurant au tableau 7.1.

La taille des écarts au niveau des DR est très faible par comparaison avec les écarts observés aux autres niveaux géographiques plus petits, qui sont traités dans les sections qui suivent.

B. Subdivisions de recensement (SDR)

Le tableau 7.2 résume les résultats concernant la cohérence entre les estimations-échantillon et les chiffres de population pour toutes les SDR échantillonnées au Canada dont le chiffre de population pour la caractéristique était supérieure à 50. Il porte sur les mêmes caractéristiques que le tableau 7.1. Les SDR ne sont pas toujours uniquement formées de RP entières. En outre, elles sont beaucoup plus petites en moyenne que les DR. Par conséquent, le niveau de cohérence n'est pas aussi bon pour les SDR que pour les DR. De façon générale, comme dans le cas des DR, la cohérence devenait meilleure à mesure qu'augmentaient les chiffres de population dans les SDR, et ce, pour toutes les caractéristiques. Comparativement aux écarts de 1991 entre le 10e et le 90e centile, les écarts de 1996 sont beaucoup plus faibles pour presque toutes les caractéristiques. Ceci s'avère également vrai pour l'intervalle entre le 25e et le 75e centile.

C. Secteurs de recensement (SR)

Le tableau 7.3 présente un résumé du niveau de cohérence pour tous les SR échantillonnés au Canada. Comme dans le cas des tableaux 7.1 et 7.2, ce tableau comprend uniquement les SR où les chiffres de population associés à la caractéristique étudiée étaient supérieurs à 50. Les SR ont en moyenne des populations plus importantes que les SDR. L'augmentation du niveau de cohérence au niveau des SR entre 1991 et 1996 est très marquée. En 1996, les 25^{e}

[7] Même lorsque les caractéristiques présentent une cohérence parfaite, les totalisations publiées des caractéristiques de base établies à partir des données-échantillon ne concordent pas parfaitement avec les totalisations des mêmes caractéristiques basées sur les données intégrales. Cette différence est attribuable au fait que les résidents des logements collectifs qui n'avaient pas à répondre aux questions-échantillon (voir le chapitre II, section B) sont inclus dans les totalisations faites à partir des données intégrales, mais exclus des totalisations fondées sur les données-échantillon.

et 75e centiles correspondent à zéro pour presque toutes les caractéristiques, contrairement aux intervalles obtenus en 1991. De plus, les intervalles entre le 10e et le 90e centiles sont beaucoup plus faibles en 1996.

De nouvelles spécifications ont été utilisées pour la formation des RP en 1996. Elles ont permis de réduire le nombre de cas de non-respect des limites de SR en regard de 1991. Ce changement pourrait contribuer à expliquer les améliorations notées en 1996.

D. Secteurs de dénombrement (SD)

Les SD sont les composantes des RP, et les RP sont le niveau le plus bas auquel les estimations-échantillon sont forcées à correspondre aux chiffres de population pour la plupart des caractéristiques. Les SD sont aussi les composantes de niveaux géographiques plus élevés (DR, SDR, SR, etc.) et plusieurs des RP, comme le montre le tableau 7.3, sont des composantes de ces niveaux supérieurs. Par conséquent, on ne peut pas s'attendre à ce que la cohérence au niveau des SD soit aussi bonne qu'aux niveaux géographiques plus élevés qui ont été étudiés. Le tableau 7.4 confirme que, pour la plupart des caractéristiques étudiées, dans les SD échantillonnés ayant un chiffre de population pour la caractéristique supérieur à 50, les écarts sont supérieurs à ceux observés aux niveaux géographiques examinés précédemment. C'est le cas aussi bien en 1996 qu'en 1991. Par rapport aux écarts de 1991 aux 10e et 90e centiles (et aux 25e et 75e centiles), les écarts de 1996 sont beaucoup plus faibles pour la plupart des caractéristiques étudiées. Les valeurs pour les ménages d'une personne sont cependant très semblables pour 1991 et 1996.

E. Incidence des changements apportés à la méthode de pondération en 1996

Il existe plusieurs explications possibles à l'augmentation du niveau de cohérence observé en 1996 par rapport à 1991 :

a) On a utilisé un nombre beaucoup moins grand de contraintes en 1996, de sorte que les contraintes retranchées ont été moins nombreuses, ce qui a permis d'utiliser des ensembles de contraintes très similaires dans toutes les RP. Le degré de cohérence entre l'estimation et le chiffre de population pour une caractéristique donnée est par conséquent plus stable parmi les RP.

b) En 1996, les coefficients de pondération calculés à l'aide de la méthode pouvaient être inférieurs à 1. Ce fait a également contribué à réduire le nombre de contraintes retranchées.

c) Les modifications au chapitre de la formation des RP ont entraîné une réduction du nombre de cas de non-respect des limites des régions géographiques de niveau supérieur. Certaines grandes régions géographiques sont maintenant constituées de RP plus complètes, ce qui contribue à déduire les écarts dans ces régions.

En conclusion, il semble que les changements apportés à la méthode de pondération en 1996 ont permis d'obtenir de meilleures estimations qu'en 1991 pour ce qui est des petites régions géographiques car, dans les régions de même niveau géographique, l'écart associé à une caractéristique donnée est généralement plus près de 0 en 1996 qu'en 1991.

Tableau 7.1 Centiles des écarts entre les estimations-échantillon et les chiffres de population (en pourcentage des chiffres de population) pour les DR – recensements de 1996 et de 1991

Caractéristiques étudiées	Centiles de 1996					Centiles de 1991				
	10e	25e	50e	75e	90e	10e	25e	50e	75e	90e
Personnes										
Hommes	0	0	0	0	0	0	0	0	0	0
Femmes	0	0	0	0	0	0	0	0	0	0
Population totale	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Âge 0 à 4	-0,44	0	0	0	0,05	-2,86	-1,07	0	1,06	2,38
Âge 5 à 9	-0,10	0	0	0	0,21	-2,14	-1,05	0	0,63	2,06
Âge 10 à 14	0	0	0	0	0,04	-1,80	-0,56	0	0,99	2,36
Âge 15 à 19	0	0	0	0	0,43	-1,87	0,51	0,55	1,76	3,15
Âge 20 à 24	-0,89	0	0	0,09	1,17	-3,71	-0,95	0,32	2,35	4,14
Âge 25 à 29	-0,95	0	0	0	1,32	-3,07	-1,40	-0,20	0,20	1,78
Âge 30 à 34	0	0	0	0	0	-1,67	-0,39	0	0,62	2,11
Âge 35 à 39	0	0	0	0	0	-3,14	-0,84	0	0,59	2,26
Âge 40 à 44	0	0	0	0	0	-2,33	-0,69	0	1,02	3,04
Âge 45 à 49	0	0	0	0	0	-2,95	-1,14	0	1,80	4,26
Âge 50 à 54	-0,17	0	0	0	0,21	-4,96	-2,14	0,13	2,02	5,03
Âge 55 à 59	-1,05	0	0	0	0,85	-6,13	-2,33	0	1,60	4,21
Âge 60 à 64	-1,51	-0,01	0	0,92	2,52	-3,69	-1,75	0,07	1,93	4,88
Âge 65 à 74	-0,25	0	0	0	0,06	-2,28	-1,08	0	0,59	2,03
Âge 75 et plus	-3,70	-1,94	-0,17	0	0,81	-7,87	-3,66	-1,07	0,67	4,65
Personnes célibataires	0	0	0	0	0	-0,10	0	0	0	0,12
Personnes mariées	0	0	0	0	0	0	0	0	0,08	0,33
Veufs et veuves	-1,53	-0,33	0	0,09	1,60	-4,22	-2,26	-0,55	0,57	2,27
Personnes divorcées	-1	0	0	0,34	2,04	-4,47	-1,82	-0,14	1,88	4,73
Personnes séparées	-7,29	-2,25	0	1,15	4,30	-9,33	-3,96	0,44	4,90	11,12
Union libre = oui [1]	-0,86	0	0	0	1,36	-	-	-	-	-
Ménages										
Ménages d'une personne	-0,57	-0,36	-0,20	-0,10	0,10	-2,54	-1,44	-0,60	0,03	0,73
Ménages de deux personnes	-0,10	0	0	0	0,01	-0,51	0	0	0,22	0,87
Ménages de trois personnes	0	0	0	0	0,10	-2,43	-0,92	0,21	1,23	3,17
Ménages de quatre personnes	0	0	0	0	0,16	-1,66	-0,57	0,04	0,92	2,15
Ménages de cinq personnes	-0,59	0	0,60	1,81	3,09	-3,83	-0,59	1,56	4,19	6,79
Ménages de six personnes ou plus	-6,55	-3,39	-0,84	0,99	2,65	-14,50	-8,69	-4,20	0,18	5,21

[1] Comme la caractéristique « Union libre = oui » n'a pas été utilisée comme contrainte en 1991, il est impossible d'établir une comparaison entre 1996 et 1991.

Tableau 7.2 Centiles des écarts entre les estimations-échantillon et les chiffres de population (en pourcentage des chiffres de population) pour les SDR – recensements de 1996 et de 1991

Caractéristiques étudiées	Centiles de 1996					Centiles de 1991				
	10e	25e	50e	75e	90e	10e	25e	50e	75e	90e
Personnes										
Hommes	-4,27	-1,22	0	1,30	4,17	-5,47	-1,69	0	1,62	5,05
Femmes	-4,49	-1,33	0	1,22	4,27	-5,49	-1,83	0	1,83	5,56
Population totale	**-2,11**	**0**	**0**	**0**	**1,94**	**-3,36**	**-0,37**	**0**	**0,26**	**3,28**
Âge 0 à 4	-16,40	-4,20	0	3,53	15,47	-19,40	-7,14	0	6,23	18,50
Âge 5 à 9	-13,20	-3,45	0	3,66	13,98	-16,50	-6,11	0	5,73	17,01
Âge 10 à 14	-13,00	-4,19	0	3,91	13,02	-17,00	-6,15	0	7,06	18,01
Âge 15 à 19	-12,90	-3,20	0	3,87	13,64	-16,80	-6,03	0	6,92	19,04
Âge 20 à 24	-14,10	-3,97	0	3,47	14,06	-21,00	-7,71	0	7,94	21,04
Âge 25 à 29	-14,20	-3,27	0	4,03	15,15	-20,30	-7,54	0	5,96	19,29
Âge 30 à 34	-12,70	-3,40	0	3,99	12,94	-17,90	-6,21	0	6,02	17,39
Âge 35 à 39	-13,00	-3,43	0	3,70	13,23	-19,20	-6,66	0	5,86	18,71
Âge 40 à 44	-13,50	-3,70	0	3,49	12,81	-19,20	-6,43	0	7,42	19,92
Âge 45 à 49	-13,00	-3,27	0	3,90	13,81	-19,20	-7,37	0	8,45	22,15
Âge 50 à 54	-14,60	-3,46	0	3,27	13,84	-23,10	-9,06	0	8,62	21,29
Âge 55 à 59	-14,80	-2,95	0	4,00	15,20	-22,30	-8,28	0	8,34	22,23
Âge 60 à 64	-15,70	-3,73	0	4,79	15,54	-22,30	-9,43	0	9,14	22,91
Âge 65 à 74	-14,80	-4,25	0	3,91	13,33	-21,20	-8,49	0	6,72	19,36
Âge 75 et plus	-18,40	-5,91	0	4,14	15,11	-26,20	-12,00	-1,31	7,74	22,02
Personnes célibataires	-5,46	-1,61	0	1,83	5,46	-7,26	-2,34	0	2,33	7,14
Personnes mariées	-5,83	-1,83	0	1,73	5,64	-6,33	-1,99	0	2,31	6,98
Veufs et veuves	-14,40	-4,13	0	3,55	14,45	-18,90	-8,35	0	6,03	17,18
Personnes divorcées	-14,90	-2,73	0	3,20	15,10	-19,60	-7,28	0	7,72	19,43
Personnes séparées	-11,10	-0,82	0	0,88	10,51	-20,50	-7,85	0	8,55	22,51
Union libre = oui	-17,40	-4,84	0	3,96	16,12	-	-	-	-	-
Ménages										
Ménages d'une personne	-11,10	-3,16	-0,14	2,87	10,48	-11,90	-5,07	-0,54	3,51	10,46
Ménages de deux personnes	-10,80	-3,29	0	3,15	10,09	-11,10	-3,71	0	4,02	11,56
Ménages de trois personnes	-10,60	-2,56	0	2,82	11,37	-15,30	-5,54	0	6,18	17,21
Ménages de quatre personnes	-10,20	-2,93	0	2,60	10,11	-14,80	-5,03	0	4,44	13,50
Ménages de cinq personnes	-7,89	-0,91	0	3,69	11,42	-14,00	-4,94	0,40	8,11	19,45
Ménages de six personnes ou plus	-9,36	-3,29	-0,28	2,12	5,84	-20,80	-10,40	-3,38	4,13	12,07

Tableau 7.3 Centiles des écarts entre les estimations-échantillon et les chiffres de population (en pourcentage des chiffres de population) pour les SR – recensements de 1996 et de 1991

Caractéristiques étudiées	Centiles de 1996					Centiles de 1991				
	10e	25e	50e	75e	90e	10e	25e	50e	75e	90e
Personnes										
Hommes	-0,20	0	0	0	0,26	-0,90	0	0	0	0,87
Femmes	-0,20	0	0	0	0,25	-0,90	0	0	0	0,79
Population totale	**0**	**0**	**0**	**0**	**0**	**-0,20**	**0**	**0**	**0**	**0,24**
Âge 0 à 4	-1,40	0	0	0	1,80	-11,00	-1,20	0	1,03	9,77
Âge 5 à 9	-1,90	0	0	0	1,72	-11,00	-1,60	0	0,94	10,20
Âge 10 à 14	-1,90	0	0	0	1,72	-11,00	-1,30	0	2,23	13,00
Âge 15 à 19	-3,10	0	0	0	3,62	-11,00	-2,70	0	3,93	11,60
Âge 20 à 24	-2,10	0	0	0	2,44	-10,00	-1,50	0	1,90	10,80
Âge 25 à 29	-1,90	0	0	0	1,89	-10,00	-1,60	0	0,28	7,31
Âge 30 à 34	-1,10	0	0	0	1,28	-7,80	-0,60	0	0,29	7,46
Âge 35 à 39	-1,00	0	0	0	1,16	-9,30	-1,30	0	0	7,88
Âge 40 à 44	-1,20	0	0	0	1,55	-10,00	-1,40	0	1,02	9,05
Âge 45 à 49	-1,10	0	0	0	1,64	-13,00	-3,10	0	3,06	12,00
Âge 50 à 54	-1,80	0	0	0	2,48	-15,00	-5,00	0	5,84	16,00
Âge 55 à 59	-3,60	0	0	0	3,62	-16,00	-5,00	0	5,84	17,30
Âge 60 à 64	-7,00	0	0	0	9,89	-17,00	-6,50	0	6,54	18,60
Âge 65 à 74	-2,70	0	0	0	2,34	-14,00	-3,00	0	1,61	12,40
Âge 75 et plus	-14,00	0	0	0	7,31	-22,00	-8,90	0	8,23	21,00
Personnes célibataires	-0,30	0	0	0	0,31	-1,20	0	0	0	1,27
Personnes mariées	-0,40	0	0	0	0,35	-1,40	0	0	0	1,59
Veufs et veuves	-4,40	0	0	0	3,36	-15,00	-5,30	0	3,87	14,90
Personnes divorcées	-2,20	0	0	0	2,97	-16,00	-5,20	0	4,08	14,20
Personnes séparées	-4,60	0	0	0	5,53	-24,00	-8,80	0	8,86	25,50
Union libre = oui	-2,50	0	0	0	3,39	-	-	-	-	-
Ménages										
Ménages d'une personne	-2,80	-0,80	0	0,29	2,06	-7,60	-3,20	-0,40	1,38	5,39
Ménages de deux personnes	-1,50	0	0	0	1,11	-3,90	-0,30	0	0,40	4,23
Ménages de trois personnes	-1,60	0	0	0	1,79	-8,90	-2,30	0	3,32	9,95
Ménages de quatre personnes	-0,90	0	0	0	2,01	-7,60	-0,90	0	2,19	9,90
Ménages de cinq personnes	-3,60	0	0	0,12	8,01	-16,00	-4,80	0	7,47	18,50
Ménages de six personnes ou plus	-6,00	-2,30	1,04	3,78	7,23	-24,00	-11,00	-0,50	7,11	16,30

Tableau 7.4 Centiles des écarts entre les estimations-échantillon et les chiffres de population (en pourcentage des chiffres de population) pour les SD – recensements de 1996 et de 1991

Caractéristiques étudiées	Centiles de 1996					Centiles de 1991				
	10e	25e	50e	75e	90e	10e	25e	50e	75e	90e
Personnes										
Hommes	-4,97	-2,05	0,02	2,14	4,80	-6,25	-2,59	0,07	2,64	6,06
Femmes	-4,91	-2,08	0	1,96	4,77	-6,29	-2,66	-0,10	2,55	6,24
Population totale	**-1,94**	**0**	**0**	**0**	**1,81**	**-3,56**	**0**	**0**	**0**	**3,35**
Âge 0 à 4	-21,40	-10,30	0	10,60	21,78	-26,40	-12,70	-0,22	12,50	26,49
Âge 5 à 9	-17,30	-8,68	-0,10	8,63	17,87	-24,20	-11,50	0,03	11,23	24,00
Âge 10 à 14	-19,20	-9,33	-0,15	9,63	20,01	-25,40	-12,20	0,07	12,23	25,28
Âge 15 à 19	-16,70	-8,26	0,11	8,94	17,86	-25,40	-12,10	0,37	13,15	27,12
Âge 20 à 24	-18,30	-9,30	-0,12	9,39	19,20	-27,90	-13,60	-0,23	13,51	28,48
Âge 25 à 29	-18,40	-9,44	-0,32	9,42	19,36	-26,10	-12,60	-0,51	11,45	25,23
Âge 30 à 34	-17,10	-8,73	-0,14	8,56	17,72	-24,40	-11,60	-0,23	11,43	24,63
Âge 35 à 39	-16,90	-8,53	-0,30	8,22	17,14	-25,10	-12,40	-0,68	11,28	24,60
Âge 40 à 44	-17,20	-9,02	-0,34	8,52	17,92	-26,30	-12,80	-0,37	11,93	26,53
Âge 45 à 49	-17,30	-9,02	-0,40	8,60	18,39	-27,40	-13,10	0	13,65	29,16
Âge 50 à 54	-19,00	-9,59	-0,54	9,37	19,54	-30,00	-15,20	-0,10	14,85	31,77
Âge 55 à 59	-19,90	-10,60	-0,74	9,96	20,66	-29,60	-15,00	0,29	15,76	31,40
Âge 60 à 64	-20,50	-10,30	0,06	10,82	22,07	-30,30	-15,40	-0,50	16,01	31,98
Âge 65 à 74	-17,30	-9,12	-0,40	8,60	17,73	-26,60	-12,60	-0,63	11,72	27,41
Âge 75 et plus	-18,20	-8,72	-0,06	8,78	18,10	-27,50	-12,70	-0,32	12,50	27,83
Personnes célibataires	-6,74	-2,76	-0,39	2,84	6,22	-8,45	-3,43	0,04	3,42	8,03
Personnes mariées	-7,85	-3,39	0,06	3,41	8,08	-8,78	-3,55	0,07	3,76	8,92
Veufs et veuves	-18,00	-9,39	0,10	8,69	17,55	-23,50	-11,50	-0,60	10,68	22,99
Personnes divorcées	-18,90	-9,61	0	9,86	19,77	-29,70	-15,20	-0,93	14,50	29,84
Personnes séparées	-33,10	-20,80	-0,36	11,74	29,78	-28,90	-11,70	3,21	17,38	32,19
Union libre = oui	-6,00	0	0	0	6,66	-	-	-	-	-
Ménages										
Ménages d'une personne	-13,10	-6,01	0,01	6,00	12,62	-13,50	-6,67	-0,35	5,69	12,38
Ménages de deux personnes	-13,80	-6,96	-0,23	6,61	13,67	-14,40	-7,08	0	7,06	15,11
Ménages de trois personnes	-16,00	-8,40	-0,17	7,88	16,21	-22,30	-11,40	-0,12	11,25	23,48
Ménages de quatre personnes	-13,90	-7,04	-0,19	7,07	14,20	-17,60	-8,41	0	8,49	18,45
Ménages de cinq personnes	-16,40	-8,23	-0,10	8,46	16,68	-22,20	-11,00	1,09	13,92	26,33
Ménages de six personnes ou plus	-13,30	-5,98	2,01	10,03	18,95	-22,60	-9,30	2,48	14,81	26,63

VIII. Variance d'échantillonnage

L'erreur d'échantillonnage a deux composantes : la variance et le biais. La variance mesure la variabilité de l'estimation par rapport à l'estimation moyenne qui résulterait de répétitions hypothétiques du processus d'enquête. Le biais est la différence entre la valeur moyenne d'une estimation qui résulterait de répétitions hypothétiques du processus d'enquête, et la valeur vraie de la variable estimée. L'erreur quadratique moyenne (EQM) mesure la variabilité de l'estimation par rapport à la valeur vraie, pour des répétitions hypothétiques du processus d'enquête. On peut montrer que l'EQM est égale à la somme de la variance et du carré du biais. L'EQM est la mesure la plus exacte de l'ampleur moyenne de l'écart entre l'estimation et la valeur vraie de la population. Si le biais est faible par rapport à la variance, la variance est une bonne approximation de l'EQM. On constate toutefois qu'en réalité, le biais s'accumule à mesure qu'on produit des estimations du recensement visant des régions géographiques plus grandes. Le biais peut donc être non significatif pour de petites régions géographiques, mais devenir élevé par rapport à la variance pour de grandes régions géographiques. Il est possible, par conséquent, que la variance soit beaucoup plus faible que l'EQM pour de grandes régions géographiques. On peut estimer la variance d'une estimation à partir de l'échantillon, mais on ne peut pas faire de même dans le cas du biais d'une estimation. Autrement dit, il n'est pas possible d'estimer avec exactitude l'EQM à partir de l'échantillon, à moins que le biais soit faible par rapport à la variance.

Dans les recensements précédents, une étude visant à produire des estimations de la variance d'échantillonnage a été réalisée. La section A donne quelques résultats de l'étude de 1986 (pour de plus amples renseignements, voir le *Guide à l'intention des utilisateurs sur la qualité des données du recensement de 1986 : Échantillonnage et pondération*). Comme au niveau des grandes régions géographiques, le biais semble être l'élément dominant de l'EQM (voir le chapitre V), le calcul de la variance d'échantillonnage ne donne pas une estimation exacte de l'EQM pour les grandes régions géographiques; c'est pourquoi il a été décidé de ne pas répéter cette étude pour le recensement de 1996. On trouvera à la section B, cependant, quelques observations relatives à l'impact qu'a eu la méthode d'estimation utilisée dans le recensement de 1996 sur la variance d'échantillonnage, par rapport au recensement de 1986.

A. Étude sur la variance d'échantillonnage du recensement de 1986

Le chapitre V présente les résultats de l'Examen du biais d'échantillonnage et décrit la nature et l'importance du biais dans l'échantillon avant la pondération. Les chapitres VI et VII présentent les résultats ayant trait au biais d'échantillonnage après la pondération. Même avec une méthode d'échantillonnage totalement dépourvue de biais, les résultats seraient sujets à la variance du simple fait que les estimations ont été produites à partir d'un échantillon. La variance peut être estimée à l'aide des données recueillies dans l'enquête-échantillon[8]. L'Étude de la variance d'échantillonnage de 1986 a été effectuée afin d'estimer l'effet des méthodes d'échantillonnage et d'estimation sur les chiffres du recensement établis à partir de données-échantillon.

Statistique Canada produit des milliers de tableaux à partir des données-échantillon (c'est-à-dire celles recueillies sur la formule 2B). Théoriquement, on pourrait fournir une mesure de précision (à savoir l'estimation de la variance d'échantillonnage) pour chaque estimation présentée dans ces tableaux. Cette mesure tiendrait compte à la fois du plan de sondage et de la méthode d'estimation. En pratique cependant, il est impossible de calculer une telle mesure pour toutes les estimations du recensement en raison du coût élevé du traitement des données. On calcule donc la variance d'échantillonnage pour un sous-ensemble seulement des estimations du recensement. Cela permet ensuite

[8] Malheureusement, la variance d'échantillonnage ne nous fournit aucune indication de l'importance de l'erreur non due à l'échantillonnage.

d'évaluer les effets combinés du plan de sondage et de la méthode d'estimation sur la variance d'échantillonnage. On peut ensuite ajuster les estimations simples de la variance d'échantillonnage, qui sont peu coûteuses à produire, pour tenir compte de cet effet et produire une estimation de la variance d'échantillonnage pour n'importe quelle estimation du recensement.

Le tableau 8.1 donne des valeurs non ajustées (simples) de l'erreur-type des estimations-échantillon du recensement. Les valeurs présentées dans ce tableau ont été établies en supposant qu'on a utilisé un échantillon aléatoire simple avec fraction de sondage de 1 sur 5, et une pondération simple de 5. Les erreurs-types sont présentées au tableau 8.1 en fonction à la fois de la taille de l'estimation du recensement et de la taille de la région géographique. Par exemple, pour une estimation de 50 personnes dans une région géographique comptant 500 personnes, l'erreur-type non ajustée est de 15.

Le tableau 8.1 contient les erreurs-types associées à un nombre restreint de valeurs du « total estimé » et du nombre total de personnes, de ménages, de logements ou de familles dans la région. La formule qui suit peut servir à calculer l'erreur-type non ajustée de n'importe quel total estimé, pour une région de n'importe quelle taille :

$$\text{ETNA} = \sqrt{\frac{4E\,(N-E)}{N}} \qquad (4)$$

où ETNA est l'erreur-type non ajustée, E est le total estimé et N est le nombre total de personnes, de ménages, de logements ou de familles dans la région. Par exemple, pour un total estimé de 750 personnes dans une régioncomptant 9 000 personnes, l'erreur-type non ajustée serait de:

$$\sqrt{\frac{4(750)(9{,}000-750)}{9{,}000}} = 52$$

L'Étude sur la variance d'échantillonnage de 1986 fournit des facteurs d'ajustement[9] par lesquels on doit multiplier les erreurs-types non ajustées pour tenir compte des effets combinés du plan de sondage et de la méthode d'estimation. Pour calculer ces facteurs d'ajustement, un échantillon de 401 RP (sur un total de 5 941 RP) a été tiré. On a réparti l'échantillon parmi les 10 provinces[10] de manière à obtenir de bonnes estimations de la variance d'échantillonnage au niveau provincial sans pour autant compromettre indûment la qualité des estimations au niveau national. Pour chaque RP dans l'échantillon, les estimations de la variance d'échantillonnage des **estimations produites par la MIQ** ont été calculées pour différentes catégories de toutes les caractéristiques énumérées au **tableau 9 du *Guide à l'intention des utilisateurs du recensement de 1986*** : ce tableau est inclus dans l'annexe D. (**NOTA IMPORTANT** : Ces facteurs ont été calculés à partir des données du recensement de 1996, alors que l'on utilisait la méthode itérative du quotient; si l'on utilise ce tableau, on suppose que l'incidence des ajustements est la même pour le recensement de 1996, même si une méthode d'estimation différente a été utilisée.) Les estimations de la variance d'échantillonnage aux niveaux provincial et national ont été obtenues en pondérant les estimations au niveau de la RP. Les facteurs d'ajustement pour chaque sous-catégorie de chacune des caractéristiques ont été calculés en divisant la racine carrée de ces estimations par l'erreur-type non ajustée correspondante. Des facteurs d'ajustement ont été calculés aux niveaux provincial et national pour chaque caractéristique en faisant la moyenne des facteurs d'ajustement de toutes les catégories. Pour de plus amples renseignements sur la façon dont ces facteurs d'ajustement ont été calculés, voir Béland, 1990.

---

9 On appelle communément « effet du plan » le carré des facteurs d'ajustement.

10 Le territoire du Yukon et les Territoires du Nord-Ouest ont été groupés avec la Colombie-Britannique.

Pour estimer l'erreur-type d'une estimation-échantillon donnée, le facteur d'ajustement s'appliquant à la caractéristique en question a été déterminé d'après l'annexe D. Le facteur d'ajustement au niveau national ou provincial pour les caractéristiques-échantillon se situait généralement entre 0,40 et 1,60. Ce facteur a ensuite été multiplié par l'erreur-type non ajustée choisie à partir du tableau 8.1.

L'exemple qui suit illustre la façon de calculer l'erreur-type ajustée. Supposons que l'estimation étudiée soit la population immigrante en Ontario. L'estimation du recensement de 1986 pour cette caractéristique était 2 081 200. Le chiffre de population obtenu pour l'Ontario au recensement de 1986 était 9 001 170. Si l'on utilise l'équation (4), qui permet de calculer l'erreur-type non ajustée, on obtient une estimation de 2 530. D'après l'annexe D, le facteur d'ajustement au niveau provincial pour la caractéristique « immigrant » est 1,12. Il s'ensuit que l'erreur-type ajustée pour cette estimation est 2 530 x 1,12 = 2 834.

Un deuxième exemple, toutefois, soulève un doute quant à l'exactitude de ces erreurs-types ajustées à titre d'estimations de la racine carrée de l'EQM. L'estimation, dans le recensement de 1986, du nombre de personnes avec « État matrimonial – Marié(e) » vivant dans des logements privés dans les SD échantillonnés était 11 771 126. Le nombre de personnes, lors du recensement de 1986, qui vivaient dans des logements privés dans des SD échantillonnés était 24 369 559. L'application de l'équation (4) donne une erreur-type non ajustée de 4 934. D'après l'annexe D, le facteur d'ajustement au niveau national pour la caractéristique « Marié(e) » est 0,25. Par conséquent, l'erreur-type ajustée pour cette estimation est 4 934 x 0,25 = 1 233. Toutefois, puisque l'état matrimonial est une caractéristique de base, on sait que le chiffre de population relatif au nombre de personnes mariées lors du recensement de 1986 qui vivaient dans des logements privés dans des SD échantillonnés était 11 778 842. La différence entre l'estimation et le chiffre de population est -7 716. Le quotient entre cette différence et l'erreur-type ajustée est -7 716/1 233 = -6,25. Un intervalle de confiance à 95 % pour une estimation serait normalement défini comme étant plus ou moins deux fois l'erreur-type ajustée. Le fait que le quotient entre la différence et l'erreur-type soit de -6,25 laisse entendre que l'erreur-type ajustée de 1 233 est une sous-estimation de la racine carrée de l'EQM.

B. Variance d'échantillonnage et biais dans l'estimation généralisée des moindres carrés

Dans Bankier, Rathwell et Majkowski, 1992, les coefficients de variation (CV) de la PEGMC pour certaines caractéristiques-échantillon ont été comparés aux CV correspondants de la MIQ. Dans les deux cas, les données du recensement de 1986 ont été utilisées. Le CV d'une estimation est la racine carrée de la variance estimée exprimée en pourcentage de l'estimation. Pour 79 RP, on a estimé les CV pour les estimations de 507 caractéristiques-échantillon au niveau des SD et 642 caractéristiques-échantillon au niveau des RP (chacune d'entre elles s'appliquant à un total minimum estimé de 60 ménages dans la population). Chacune des estimations au niveau des RP et au niveau des SD a été classée comme estimation faible (inférieure ou égale à la médiane des estimations) ou comme estimation élevée (supérieure à la médiane des estimations). On a constaté que la valeur médiane des CV pour les estimations élevées au niveau des RP était de 5 % pour la PEGMC, tandis qu'elle était de 6 % pour la MIQ. La valeur médiane des CV pour les estimations faibles au niveau des RP était de 13 % pour la PEGMC et de 15 % pour la MIQ. La valeur médiane des CV pour les estimations élevées au niveau des SD était de 10 % pour la PEGMC et de 12,5 % pour la MIQ. La PEGMC a donc permis une réduction des CV par rapport à la MIQ tant au niveau des SD qu'à celui des RP. Puisque les variations à des niveaux géographiques supérieurs sont simplement la somme des variances au niveau des RP, ces réductions des CV devraient se répercuter sur les niveaux géographiques plus élevés.

Il a été indiqué au chapitre V que l'échantillon du recensement affichait des biais faibles, mais significatifs. Ces biais ne sont pas significatifs par rapport à la variance d'échantillonnage au niveau des RP. À des niveaux géographiques supérieurs, toutefois, le biais touchant une caractéristique peut s'accumuler s'il produit systématiquement des

surestimations ou des sous-estimations. Il semble que l'effet du biais soit plus important dans le cas de la PEGMC que dans celui de la MIQ. On peut le voir au tableau 6.5 du chapitre VI, où la PEGMC présente des différences population/ estimation plus faibles que la MIQ pour les plus petites régions géographiques. Cependant, à mesure que la région géographique grossit, l'amélioration observée n'est plus aussi importante, et en 1991 la méthode itérative du quotient donnait des résultats supérieurs à ceux de la PEGMC au niveau provincial. Outre le biais introduit par l'échantillonnage et le traitement, Bankier, Rathwell et Majkowski,1992, montrent, par une étude de Monte Carlo, que l'estimateur PEGMC est lui-même biaisé, bien que le biais relatif soit inférieur à 1 % pour 50 % des caractéristiques étudiées. Fait plus important, toutefois, la variance estimée des estimateurs PEGMC affiche un biais relatif médian de -25 % au niveau des RP. Ces estimateurs ont donc tendance à sous-estimer la variance vraie. Les estimateurs MIQ souffrent peut-être de biais semblables, mais aucune étude n'a été faite à ce sujet.

Tableau 8.1 Estimations non ajustées des erreurs-types des estimations-échantillon

Total estimé	Nombre total estimatif de personnes, de ménages et de logements								
dans la région	500	1 000	2 500	5 000	10 000	25 000	50 000	100 000	250 000
50	15	15	15	15	15	15	15	15	15
100	18	19	20	20	20	20	20	20	20
250	22	25	30	30	30	30	30	30	30
500	0	30	40	40	45	45	45	45	45
1 000		0	50	55	60	60	65	65	65
2 500			0	70	85	95	95	100	100
5 000				0	100	130	130	140	140
10 000					0	150	180	190	200
25 000						0	220	270	300
50 000							0	320	400
100 000								0	490
250 000									0

Total estimé	Nombre total estimatif de personnes, de ménages et de logements					
dans la région	500 000	1 000 000	2 500 000	5 000 000	10 000 000	25 000 000
50	15	15	15	15	15	15
100	20	20	20	20	20	20
250	30	30	30	30	30	30
500	45	45	45	45	45	45
1 000	65	65	65	65	65	65
2 500	100	100	100	100	100	100
5 000	140	140	140	140	140	140
10 000	200	200	200	200	200	200
25 000	310	310	310	320	320	320
50 000	420	440	440	440	450	450
100 000	570	600	620	630	630	630
250,000	710	870	950	970	990	990
500 000	0	1 000	1 260	1 340	1 380	1 400
1 000,000		0	1 550	1 790	1 900	1 960
2 500 000			0	2 240	2 740	3 000
5 000 000				0	3 160	4 000
10 000 000					0	4 900

IX. Conclusion

L'échantillonnage fait maintenant partie intégrante du processus de recensement. Son utilisation permet de faire des économies substantielles et de réduire de façon appréciable le fardeau du répondant, ou encore, d'élargir la portée d'un recensement sans frais supplémentaires. Le prix de ces avantages est l'introduction d'une erreur d'échantillonnage dans les estimations obtenues à partir de l'échantillon. Cet effet se fait particulièrement sentir sur les estimations de faible valeur du recensement, qu'il s'agisse des chiffres se rapportant à des catégories peu usuelles au niveau provincial ou provincial, ou de ceux qui correspondent à des régions géographiques de petite taille. Il convient de souligner que les erreurs de réponse et les erreurs de dépouillement contribuent également à l'erreur totale des estimations du recensement et, là encore, ce sont les estimations de faible valeur qui sont particulièrement sensibles aux effets de ces erreurs non dues à l'échantillonnage. Par conséquent, bon nombre de ces faibles valeurs n'auraient qu'une fiabilité relative même si elles étaient fondées sur un recensement intégral. Pour ce qui est du recensement de 1996, les estimations inférieures ou égales à 50 établies à partir des données-échantillon sont généralement très peu fiables, tandis que les estimations allant jusqu'à 500 tendent à avoir des erreurs-types de plus de 10 % de leur valeur.

Pour de nombreuses caractéristiques, un certain biais a été détecté dans l'échantillon. Il a été établi que ce biais avait été causé partiellement par le traitement des données et à l'étape du contrôle et de l'imputation. Le reste du biais provient sans doute d'un ou plusieurs facteurs tels que le biais de non-réponse, le biais de réponse ou la sélection par les recenseurs d'un échantillon biaisé. Les méthodes utilisées pour pondérer les données-échantillon en vue d'obtenir des estimations à l'échelle de la population totale ont été appliquées avec succès et, de façon générale, ont permis d'atteindre les niveaux prévus de cohérences des estimations-échantillon et des chiffres de population. Bien que la cohérence qui a été obtenue aux niveaux provincial et national soit meilleure qu'en 1991, elle est quand même plus faible que ce à quoi on aurait pu s'attendre, compte tenu de la cohérence améliorée atteinte lors des tests pour les plus petits niveaux géographiques. Cela s'explique probablement par les petits biais accumulés dans l'échantillon sur de nombreuses années.

La méthode d'estimation du recensement sera réévaluée en vue du recensement de 2001, afin de voir s'il est possible d'améliorer la cohérence entre les estimations-échantillon et les chiffres de population au niveau des provinces et du Canada, tout en maintenant une bonne cohérence au niveau des SD. On devrait pouvoir obtenir en même temps des estimations plus fiables de l'erreur quadratique moyenne des estimations du recensement.

ANNEXES

Annexe A – Glossaire

Les définitions des termes, variables et concepts du recensement sont présentées ici telles qu'elles figurent dans le *Dictionnaire du recensement de 1996* (n° 92-351-XPF au catalogue). L'utilisateur doit se reporter au *Dictionnaire du recensement de 1996* pour obtenir les définitions complètes et des observations additionnelles se rapportant aux concepts, comme des renseignements sur les variables directes et dérivées, ainsi que leur univers respectif.

Division de recensement (DR) : Terme générique qui désigne les régions géographiques établies en vertu de lois provinciales et qui sont des régions géographiques intermédiaires entre la municipalité (subdivision de recensement) et la province. Il peut s'agir de comtés, de districts régionaux, de municipalités régionales et d'autres genres de régions créées en vertu des lois provinciales.

État matrimonial : Situation conjugale d'une personne : marié(e) (y compris l'union libre); séparé(e); divorcé(e); veuf ou veuve; célibataire (jamais marié[e]).

Logement privé : Ensemble distinct de pièces d'habitation ayant une entrée privée donnant sur l'extérieur ou sur un corridor, un hall, un vestibule ou un escalier commun à l'intérieur. L'entrée doit donner accès au logement sans que l'on ait à passer par les pièces d'habitation de quelqu'un d'autre.

Logement privé occupé : Logement privé occupé de façon permanente par une personne ou un groupe de personnes. Sont également inclus dans cette catégorie les logements privés dont les résidents habituels sont temporairement absents le jour du recensement.

Ménage : Personne ou groupe de personnes (autres que des résidents étrangers) occupant un même logement et n'ayant pas de domicile habituel ailleurs au Canada. Il peut se composer d'un groupe familial (famille de recensement), avec ou sans autres personnes hors famille de recensement, de deux familles ou plus partageant le même logement, d'un groupe de personnes non apparentées ou d'une personne seule.

Ménage privé : Personne ou groupe de personnes (autres que des résidents étrangers) occupant un logement privé et n'ayant pas de domicile habituel ailleurs au Canada.

Secteur de dénombrement (SD) : Correspond à la région géographique dénombrée par un recenseur. Il s'agit de la plus petite région géographique normalisée pour laquelle des données du recensement sont établies. Tout le territoire du Canada est divisé en SD.

Secteur de recensement (SR) : Petites unités géographiques représentant des collectivités urbaines ou rurales créées au sein de régions métropolitaines de recensement et d'agglomérations de recensement dont le noyau urbanisé compte 50 000 habitants ou plus d'après le recensement précédent.

Subdivison de recensement (SDR) : Terme générique qui désigne les municipalités (telles que définies par les lois provinciales) ou leurs équivalents (par exemple, les réserves indiennes, les établissements indiens et les territoires non organisés).

Annexe B – Contraintes au niveau des RP et des SD appliquées aux coefficients de pondération du recensement de 1996

(**Nota** : Les contraintes ajoutées lors du deuxième exercice effectué en 1996 sont indiquées par le symbole « # ».)

Contraintes au niveau des RP (personnes)

– Total des personnes
– Total des personnes âgées de ≥ 15 ans

– Hommes
– Hommes âgés de ≥ 15 ans

– Personnes âgées de 0 à 4 ans
– Personnes âgées de 5 à 9 ans
– Personnes âgées de 10 à 14 ans
– Personnes âgées de 15 à 19 ans
– Personnes âgées de 20 à 24 ans
– Personnes âgées de 25 à 29 ans
– Personnes âgées de 30 à 34 ans
– Personnes âgées de 35 à 39 ans
– Personnes âgées de 40 à 44 ans
– Personnes âgées de 45 à 49 ans
– Personnes âgées de 50 à 54 ans
– Personnes âgées de 55 à 59 ans
– Personnes âgées de 60 à 64 ans #
– Personnes âgées de 65 à 74 ans
– Personnes âgées de ≥ 75 ans

– Personnes mariées
– Personnes célibataires
– Personnes divorcées
– Veufs et veuves
– Union libre #

Contraintes au niveau des SD

– Total des ménages dans le SD
– Total des personnes dans le SD

Contraintes au niveau des RP (ménages)

– Ménages d'une personne #
– Ménages de deux personnes
– Ménages de trois personnes
– Ménages de quatre personnes
– Ménages de cinq personnes

Annexe C – Renseignements additionnels sur les statistiques utilisées dans l'Examen du biais d'échantillonnage

Que X représente la valeur connue d'une caractéristique 2A au niveau de la division de recensement (DR) et que $\hat{X}^{(0)}$ représente l'estimateur de Horvitz-Thompson de X. $\hat{X}^{(0)}$ a été calculé en multipliant le total-échantillon non pondéré pour la caractéristique de chaque SD échantillonné par l'inverse de la fraction de sondage des ménages réalisée pour le SD, puis en faisant la somme des résultats au niveau de la DR. Les SD non échantillonnés ont été exclus de l'analyse. L'écart-type de $\hat{X}^{(0)}$, std $(\hat{X}^{(0)}) = \sqrt{V(\hat{X}^{(0)})}$ a été calculé selon l'hypothèse que des échantillons aléatoires simples de ménages ont été tirés indépendamment dans chaque SD (en fait, des échantillons aléatoires systématiques indépendants ont été tirés). Les variances ont donc été calculées au niveau des SD, puis on en a fait la somme au niveau de la DR. Les valeurs S^2 de la population ont été utilisées dans les calculs de la variance. Voir Cochran, 1977, pages 23-24, pour les formules de variance relatives aux caractéristiques des personnes et des familles, et pages 50-52 pour les formules de variance relatives aux caractéristiques des ménages et des logements.

Puisque les valeurs $\hat{X}^{(0)}$ sont des estimateurs de Horvitz-Thompson, ils sont sans biais pour X. L'échantillonnage a été fait indépendamment dans différents SD. Par conséquent, les valeurs $\hat{X}^{(0)}$ sont la somme de *n* variables aléatoires indépendantes, où *n* est le nombre de SD dans la DR. Puisque 90 % des DR comptaient plus de 25 SD (la moyenne était de 140), *n* est très élevé dans la plupart des DR. Par conséquent, selon le théorème central limite, $Z^{(0)} = (\hat{X}^{(0)} - X)/\text{std}\,(\hat{X}^{(0)})$ devrait avoir une distribution approximativement normale (0,1) (voir Kendall et Stuart, 1963, p. 193). Ce ne serait pas le cas, toutefois, si les réponses 2B présentaient un biais important, pour quelque raison que ce soit.

Les valeurs $Z^{(0)}$ ont été produites pour l'ensemble des 281 DR échantillonnées au Canada, à l'égard des caractéristiques 2A données au chapitre V. Afin d'évaluer la normalité des valeurs $Z^{(0)}$ au niveau des DR, des histogrammes des valeurs $Z^{(0)}$ superposés à une FDP (fonction de densité de probabilité) normale ont été produits.

De plus, pour vérifier si $Z^{(0)}$ était tiré d'une distribution normale dont la moyenne est zéro (c.-à-d. si la méthode de sélection de l'échantillon était sans biais), on a calculé la moyenne $\bar{Z}^{(0)} = \sum_{i=1}^{m} Z_i^{(0)}/m$ où $m = 281$ (le nombre de DR) et $Z_i^{(0)}$ est la valeur de $Z^{(0)}$ pour la i^e DR. En outre, l'écart-type des $Z_i^{(0)}$ a été déterminé, compte tenu que $\text{std}^2(Z^{(0)}) = \sum_{i=1}^{m} (Z_i^{(0)} - \bar{Z}^{(0)})/(m-1)$. Puis la statistique T donnée par $T_z = \sqrt{m}\,\bar{Z}^{(0)}/\text{std}(Z^{(0)})$ a été calculée. Si la méthode de sélection de l'échantillon était non biaisée, T devrait suivre la distribution *t* de Student avec m-1 degrés de liberté. La probabilité que $|T_Z| > 1\ 960$ si la méthode de sélection de l'échantillon était non biaisée est inférieure à 0,05. Donc, si $|T_Z| > 1\ 960$, l'hypothèse que la méthode de sélection de l'échantillon était non biaisée sera rejetée et la différence entre l'estimation-échantillon et le chiffre de population sera dite statistiquement significative au niveau de 5 %.

Annexe D – Facteurs d'ajustement de l'erreur-type de 1986 au niveau national ou provincial et centiles des facteurs au niveau de la région de pondération

Caractéristiques	Facteur national ou provincial	Centiles des facteurs au niveau de la RP					
		1	50	75	90	95	99
Caractéristiques de la population							
Âge							
Groupes d'âge 0 à 4, 5 à 9, 10 à 14, 15 à 19, 20 à 24, 25 à 29	0,18	0,05	0,19	0,29	0,35	0,49	0,52
Groupes d'âge 30 à 34, 35 à 44, 45 à 54, 55 à 59, 60 à 64, 5+, 15+	0,36	0,13	0,33	0,46	0,51	0,56	0,61
Groupe d'âge 65+	0,00	-	-	-	-	-	-
Sexe	0,00	-	-	-	-	-	-
État matrimonial							
Célibataires, personnes mariées (sauf les personnes séparées)	0,25	0,04	0,23	0,31	0,42	0,49	0,55
Personnes séparées, divorcées, veufs et veuves	0,88	0,55	0,84	0,98	1,06	1,15	1,20
Plus haut niveau de scolarité							
Plus haut grade, certificat ou diplôme/ Total des années de scolarité	0,90	0,75	0,95	1,06	1,14	1,19	1,25
Principal domaine d'études	1,20	0,84	1,16	1,22	1,28	1,35	1,43
Mobilité							
Personnes n'ayant pas déménagé	1,21	0,83	1,23	1,27	1,32	1,36	1,41
Personnes ayant déménagé (migrants, non-migrants)	1,61	0,90	1,60	1,75	1,85	1,97	2,09

Caractéristiques	Facteur national ou provincial	Centiles des facteurs au niveau de la RP					
		1	50	75	90	95	99
Caractéristiques de la population							
Période d'immigration							
Avant 1946, 1946-1966	0,98	0,76	1,02	1,10	1,22	1,37	1,45
1967-1977, 1978-1982, 1983-1986	1,51	0,80	1,45	1,55	1,78	1,90	2,11
Âge à l'immigration	1,10	0,71	1,15	1,29	1,38	1,44	1,54
Lieu de naissance							
Né(e) au Canada	1,09	0,82	1,08	1,16	1,18	1,20	1,21
Né(e) à l'extérieur du Canada	1,35	1,11	1,34	1,43	1,60	1,67	1,75
Population immigrante/non immigrante	1,12	0,81	1,10	1,24	1,38	1,46	1,52
Citoyenneté							
Canada, par naissance	1,13	0,88	1,14	1,17	1,20	1,27	1,32
Autres	1,59	1,04	1,40	1,65	1,88	1,95	2,12
Origine ethnique							
Anglais, français	1,20	0,73	1,16	1,25	1,31	1,40	1,46
Autres	1,65	1,07	1,57	1,70	1,89	1,99	2,11
Langue parlée à la maison							
Anglais, français, anglais et français, anglais et langues non officielles	1,12	0,50	1,09	1,35	1,75	1,89	2,09
Autres groupes linguistiques	1,76	0,99	1,68	1,89	2,01	2,20	2,41
Langues officielles							
Anglais, français, anglais et français	1,05	0,69	1,01	1,18	1,31	1,42	1,58
Autres groupes linguistiques	1,49	0,90	1,50	1,68	1,76	1,79	1,91
Langue maternelle – Anglais	1,21	0,83	1,23	1,27	1,32	1,36	1,41
Terre-Neuve, Île-du-Prince-Édouard, Nouvelle-Écosse, Colombie-Britannique	0,92	0,24	0,96	1,45	1,62	1,90	2,23

Caractéristiques	Facteur national ou provincial	Centiles des facteurs au niveau de la RP					
		1	50	75	90	95	99
Caractéristiques de la population							
Québec	1,15	0,18	1,10	1,51	1,76	1,81	1,99
Autres provinces	0,45	0,12	0,48	0,71	0,96	1,12	1,38
Canada	0,53	-	-	-	-	-	-
Langue maternelle – Français							
Québec	0,42	0,14	0,45	0,52	0,61	0,76	0,91
Nouveau-Brunswick	0,75	0,19	0,79	0,98	1,24	1,60	1,84
Autres provinces	1,04	0,09	1,12	1,49	1,71	1,89	2,06
Canada	0,77	-	-	-	-	-	-
Langue maternelle – Autres groupes linguistlques	1,70	0,73	1,63	2,11	2,44	2,51	2,60
Industrie/profession	0,92	0,25	0,80	1,13	1,25	1,31	1,38
Travail en 1985	0,89	0,62	0,92	1,14	1,22	1,29	1,31
Semaines travaillées en 1985	0,94	0,68	0,99	1,18	1,29	1,33	1,39
Heures travaillées pendant la semaine de référence	0,83	0,63	0,85	1,01	1,14	1,19	1,24
Année du dernier travail							
En 1986, en 1985, avant 1985	0,89	0,60	0,94	0,99	1,05	1,11	1,20
Jamais travaillé	1,18	0,80	1,15	1,34	1,43	1,50	1,67
Catégorie de travailleurs							
Travailleurs rémunérés	0,72	0,56	0,75	0,86	0,93	0,95	0,98
Travailleurs autonomes, entreprises non constituées en société, travailleurs familiaux non rémunérés	0,93	0,68	0,96	1,08	1,13	1,15	1,18

Caractéristiques	Facteur national ou provincial	Centiles des facteurs au niveau de la RP					
		1	50	75	90	95	99
Caractéristiques de la population							
Participation à la population active							
Personnes occupées	0,75	0,59	0,76	0,83	0,86	0,91	0,93
Personnes en chômage	1,06	0,76	1,04	1,14	1,20	1,27	1,38
Personnes inactives	1,25	0,91	1,30	1,43	1,50	1,58	1,63
Principale source de revenu							
Salaires et traitements	0,65	0,42	0,67	0,80	0,85	0,87	0,92
Autres	1,05	0,71	1,00	1,12	1,17	1,20	1,24
Incapacité							
Limité(e) à la maison, à l'école et au travail	0,94	0,69	0,96	1,11	1,29	1,34	1,42
N'est pas limité(e)	0,61	0,41	0,58	0,69	0,74	0,78	0,81
Situation des particuliers dans la famille de recensement							
Conjoint, conjointe, enfant	0,20	0,05	0,20	0,24	0,26	0,28	0,31
Parent seul de sexe féminin	0,45	0,14	0,43	0,51	0,55	0,61	0,68
Parent seul de sexe masculin, personne ne faisant pas partie d'une famille de recensement	0,68	0,35	0,65	0,79	0,89	0,99	1,14
Situation des particuliers dans la famille économique							
Conjoint, conjointe	0,14	0,06	0,16	0,21	0,28	0,34	0,36
Parent seul, enfant	0,32	0,16	0,34	0,39	0,44	0,47	0,53
Autres membres de la famille	0,74	0,24	0,70	0,84	1,03	1,09	1,18
Nombre de personnes dans la famille de recensement	0,04	0,00	0,00	0,05	0,07	0,09	0,11

Caractéristiques	Facteur national ou provincial	Centiles des facteurs au niveau de la RP					
		1	50	75	90	95	99
Caractéristiques de la population							
Nombre de personnes dans la famille économique	0,18	0,08	0,19	0,24	0,33	0,41	0,45
Âge du conjoint, de la conjointe ou de la personne de référence dans la famille économique	1,42	0,80	1,37	1,53	1,60	1,78	1,91
Toutes les autres caractéristiques de la population	1,00	-	-	-	-	-	-
Caractéristiques des ménages et des logements							
Type de construction							
Logement individuel non attenant	0,33	0,05	0,35	0,55	0,67	0,75	0,89
Appartement de moins de 5 étages	0,57	0,12	0,56	0,70	0,83	0,99	1,26
Autres	0,91	0,18	0,88	0,99	1,18	1,23	1,32
Mode d'occupation	0,00	-	-	-	-	-	-
Période de construction	0,78	0,61	0,75	0,82	0,89	0,99	1,24
Principal type de chauffage/Principal combustible utilisé pour le chauffage	0,87	0,18	0,86	1,04	1,12	1,25	1,32
Chauffage central							
Avec	0,42	0,09	0,38	0,54	0,60	0,70	0,89
Sans	0,78	0,23	0,79	0,91	1,03	1,12	1,20
Taille du ménage							
Ménage d'une personne	0,00	-	-	-	-	-	-
Autres	0,76	0,19	0,72	1,09	1,17	1,21	1,30

Caractéristiques	Facteur national ou provincial	Centiles des facteurs au niveau de la RP					
		1	50	75	90	95	99
Caractéristiques des ménages et des logements							
Nombre de pièces	0,80	0,57	0,78	0,90	0,97	1,10	1,20
Âge du soutien du ménage							
25 à 34, 55 à 64, 65 à 74, 75+	0,25	0,06	0,24	0,35	0,48	0,53	0,62
0 à 24, 35 à 44, 45 à 54	0,92	0,38	0,90	1,05	1,14	1,21	1,30
Sexe du soutien du ménage							
Hommes	0,20	0,09	0,24	0,31	0,34	0,36	0,37
Femmes	0,47	0,16	0,43	0,54	0,64	0,74	0,89
Loyer brut/Loyer brut, sous forme de pourcentage du revenu du ménage	0,75	0,48	0,79	0,91	0,94	0,96	1,01
Principales dépenses de propriété/ Principales dépenses de propriété, sous forme de pourcentage du revenu du ménage	0,84	0,62	0,87	0,95	1,01	1,04	1,11
Revenu du ménage	0,75	0,51	0,73	0,82	0,90	0,95	1,03
Valeur du logement	0,90	0,67	0,91	1,00	1,05	1,12	1,18
Condominium enregistré							
Partie	0,63	0,18	0,59	0,84	0,93	1,11	1,30
Ne fait pas partie	0,15	0,07	0,14	0,19	0,28	0,39	0,47
Type de logement – Ménages unifamiliaux							
Sans personnes additionnelles	0,22	0,05	0,20	0,27	0,33	0,36	0,40
Avec personnes additionnelles	0,50	0,20	0,48	0,61	0,72	0,74	0,79

Caractéristiques	**Facteur national ou provincial**	**Centiles des facteurs au niveau de la RP**					
		1	**50**	**75**	**90**	**95**	**99**
Caractéristiques des ménages et des logements							
Type de logement – Ménages non familiaux	0,00	-	-	-	-	-	-
Type de logement – Autres	1,12	0,54	1,05	1,26	1,40	1,51	1,67
Toutes les autres caractéristiques des ménages et des logements	1,00	-	-	-	-	-	-
Caractéristiques de la famille de recensement							
Structure de la famille de recensement							
Conjoint et conjointe	0,20	0,09	0,21	0,26	0,29	0,33	0,36
Parent seul de sexe masculin	0,64	0,21	0,62	0,81	0,84	0,91	1,04
Parent seul de sexe féminin	0,46	0,19	0,45	0,57	0,65	0,69	0,74
Genre de famille de recensement							
Famille principale	0,23	0,04	0,24	0,28	0,31	0,34	0,39
Famille secondaire	0,90	0,62	0,93	1,15	1,28	1,33	1,40
Groupes d'âge des enfants à la maison	0,78	0,40	0,70	0,91	0,98	1,09	1,19
Activité du conjoint, de la conjointe ou du parent seul							
Conjoint, parent seul, conjoint et conjointe dans la population active	0,40	0,23	0,43	0,50	0,55	0,59	0,71
Conjointe dans la population active	0,61	0,41	0,60	0,68	0,74	0,78	0,82
Autres	0,72	0,30	0,68	0,80	0,90	0,99	1,12

Caractéristiques	Facteur national ou provincial	Centiles des facteurs au niveau de la RP					
		1	50	75	90	95	99
Caractéristiques de la famille de recensement							
Travail en 1985 du conjoint, de la conjointe ou du parent seul							
A travaillé en 1985	0,48	0,11	0,45	0,50	0,54	0,57	0,59
N'a pas travaillé en 1985	0,93	0,60	0,90	1,04	1,18	1,26	1,30
Toutes les autres caractéristiques de la famille de recensement	1,00	-	-	-	-	-	-
Caractéristiques de la famille économique							
Structure de la famille économique							
Familles conjoint-conjointe	0,29	0,13	0,30	0,36	0,48	0,56	0,68
N'est pas une famille conjoint-conjointe	0,56	0,35	0,50	0,66	0,81	0,90	1,06
Langue maternelle de la famille de la personne de référence – Anglais							
Terre-Neuve, Île-du-Prince-Édouard, Colombie-Britannique	0,25	0,09	0,20	0,31	0,45	0,66	0,91
Québec	0,49	0,25	0,47	0,50	0,69	0,83	1,05
Autres provinces	0,18	0,07	0,19	0,22	0,24	0,27	0,31
Canada	0,27	-	-	-	-	-	-
Langue maternelle de la famille de la personne de référence – Français							
Québec	0,12	0,04	0,13	0,17	0,21	0,29	0,36
Autres provinces	0,88	0,30	0,90	1,07	1,21	1,28	1,35
Canada	0,40	-	-	-	-	-	-

Caractéristiques	Facteur national ou provincial	Centiles des facteurs au niveau de la RP					
		1	50	75	90	95	99
Caractéristiques de la famille économique							
Langue maternelle de la famille de la personne de référence – Autre que l'anglais et le français							
Terre-Neuve, Nouvelle-Écosse	0,75	0,38	0,74	0,80	0,91	0,99	1,10
Autres provinces	0,50	0,21	0,45	0,57	0,82	0,84	0,99
Canada	0,56	-	-	-	-	-	-
Toutes les autres caractéristiques de la famille économique	1,00	-	-	-	-	-	-

Annexe E – Produits et services

Pour mettre en valeur les données du recensement, on doit les présenter de façon qu'elles soient pertinentes et accessibles pour les clients, qu'ils soient décideurs, analystes de politiques, bibliothécaires, spécialistes en commercialisation, chercheurs, étudiants, etc. Plusieurs nouveaux produits et services sont offerts à l'occasion du recensement de 1996.

1. Accessibilité accrue sur support électronique

Un nombre accru de clients ont demandé que les produits du recensement soient offerts sur support électronique utilisable sur des ordinateurs personnels. Bien qu'on ait conservé certains produits clés sur support imprimé, un plus grand nombre de données du recensement sont offertes sur CD-ROM et sur diskettes. Ces supports contiennent des logiciels de présentation et de totalisation pour environnement Windows, qui rendent les données faciles à utiliser. Pour la première fois, les clients ont pu obtenir des données gratuitement dans Internet, au site Web de Statistique Canada : http://www.statcan.ca.

2. Données régionales diffusées plus tôt

Les données du recensement à l'échelon des petites régions géographiques ont été diffusées beaucoup plus tôt que dans les années antérieures. Le jour même de chaque diffusion, les données des profils étaient accessibles à l'échelon des collectivités (subdivisions et divisions de recensement) et, un mois plus tard, les données pour des régions aussi petites que les secteurs de recensement, les secteurs de dénombrement et les régions de tri d'acheminement étaient diffusées.

3. Totalisations du recensement accessibles selon le code postal

Comme partie intégrante des produits normalisés, les tableaux sommaires de base et les profils de secteurs ont été diffusés pour les régions de tri d'acheminement, qui correspondent aux trois premiers caractères du code postal. Les données selon le code postal complet peuvent être obtenues par l'intermédiaire d'une demande de produit personnalisé, sous réserve des restrictions en matière de confidentialité.

4. Nouvelles données recueillies en 1996

Pour la première fois, des données sont diffusées sur les activités à la maison non rémunérées, le lieu de travail pour tous les échelons géographiques, le mode de transport utilisé pour se rendre au travail et les groupes de population.

5. Produits géographiques améliorés

On a non seulement amélioré la qualité de nombreuses cartes servant à la diffusion des données du recensement, mais aussi repris la production d'une série de cartes des circonscriptions électorales fédérales. Le produit *GéoSuite* (auparavant GéoRéf), un outil électronique pour environnement Windows qui permet aux utilisateurs d'explorer les liens entre différents échelons géographiques, a également été amélioré par l'ajout de listes de référence des secteurs de dénombrement.

Bibliographie

La Division des opérations du recensement a produit les parties suivantes du présent rapport : l'Introduction, les annexes A et E ainsi que les Centres régionaux de consultation.

Béland, Y. «Results and Methodology of the 1986 Sampling Variance Study». Rapport interne de Statistique Canada, 1990.

Bankier, M., S. Rathwell, et M. Majkowski. «Two-step Generalized Least Squares Estimation in the 1991 Canadian Census». Document de travail de la Direction de la méthodologie, Division des méthodes d'enquêtes sociales, 1992. Numéro 92-007E au catalogue.

Bankier, M., A.-M. Houle, et M. Luc. Calibration Estimation in the 1991 and 1996 Canadian Censuses. Proceedings of the Survey Methods Research Section, American Statistical Association (1997), pages 66-75.

Brackstone, G.J. et J.N.K. Rao. «An Investigation of Raking Ratio Estimators». Sankhya, volume 41, série C, partie 2, pages 97-114, 1979.

Bureau fédéral de la statistique. Huitième recensement du Canada, 1941, Rapport administratif du statisticien en chef, Ottawa : King's Printer, 1945.

Bureau fédéral de la statistique. Neuvième recensement du Canada, 1951, vol. XI, Rapport administratif, Ottawa : Queen's Printer, 1955.

Bureau fédéral de la statistique. Sampling in the Census. S.M.S.03.5, 1968.

Bureau fédéral de la statistique. Recensement du Canada de 1961. Revue générale, Bulletin 7.2-12, numéro 99-537 au catalogue, Ottawa : Queen's Printer, 1970.

Cochran, W. «Sampling Techniques 3rd Edition». John Wiley and Sons : Toronto, 1977.

Fellegi, I.P. «Response Variance and its Estimation». Journal of the American Statistical Association, 59, pages 1016-1041, 1964.

Hansen, M.H., W.N. Hurwitz, et M.A. Bershad. «Measurement Errors in Censuses and Surveys». Bulletin of the International Statistical Institute, 38, pages 359-374, 1959.

Kendall, M.G. et A. Stuart. «The Advanced Theory of Statistics», volume 1, Charles Griffin and Company Limited, London, 1963.

Kruszynski, G. Evaluation of the 1996 Weighting Factor. Rapport interne de la Division de la géographie, Statistique Canada, 1999.

Majkowski, M. «1991 Census 2A/2B Discrepancies». Rapport interne de Statistique Canada, 1992a.

Majkowski, M. «Investigation into Large Population/Estimate Differences in the 1991 Census». Rapport de Statistique Canada, 1994.

Royce, D. «The Use of Sampling in the 1981 Canadian Census». Rapport interne de Statistique Canada, 1983.

Sarndal, C., B. Swensson, et J. Wretman. «Model Assisted Survey Sampling». Springer-Verlag : New York, 1992.

Statistique Canada. Recensement du Canada de 1971. Revue générale, vol. VI, partie 1, numéro 99-740 au catalogue, Ottawa, 1976.

Statistique Canada. Recensement du Canada de 1976. Rapport administratif, partie 1, numéro 99-850 au catalogue, Ottawa, 1980.

Statistique Canada. Recensement du Canada de 1976. «Quality of Data, Series 1: Sources of Error - Sampling and Weighting». Numéro 99-844 au catalogue, Ottawa, 1980.

Statistique Canada. Recensement du Canada de 1981. Guide sommaire : Population totale. Numéro 99-902 au catalogue, Ottawa, 1983.

Statistique Canada. Recensement du Canada de 1986. Le Recensement en bref. Numéro 99-104F au catalogue, Ottawa, 1988.

Statistique Canada. Guide à l'intention des utilisateurs sur la qualité des données du recensement de 1986 : Échantillonnage et pondération. Numéro 99-136F au catalogue, Ottawa, 1990.

Statistique Canada. Recensement du Canada de 1991. Le recensement de 1991 en bref, numéro 92-305F au catalogue, Ottawa, 1992.

Thieverge, S. «Bias in the 1996 Census Sample». Rapport interne de la Division des méthodes d'enquêtes sociales, Statistique Canada, 1999.

Centres régionaux de consultation

Les centres régionaux de consultation de Statistique Canada sont situés un peu partout au pays et chacun possède une série complète des publication courantes et des documents de référence que les clients peuvent consulter sans frais ou acheter. On y trouve aussi des disquettes pour micro-ordinateurs, des CD-ROM, des cartes et d'autres produits et services, dont CANSIM.

Chaque centre de consultation vous offre une gamme de produits et services additionnels. D'une part, les Services de diffusion : service téléphonique gratuit de renseignements pour les données de base les plus récentes. D'autre part, les conseils : identification de vos besoins, détermination des sources ou de la disponibilité des données, consolidation et intégration de données provenant de diverses sources et élaboration de profils, analyse de faits saillants ou de tendances et, pour terminer, formation sur les produits, services et concepts de Statistique Canada ainsi que l'utilisation de données statistiques.

Pour en savoir davantage sur les services offerts dans les centres régionaux de consultation, il suffit de communiquer avec le centre de sa région, dont la liste se trouve ci-dessous. Les personnes qui habitent à l'extérieur de la zone de communication locale disposent d'un numéro sans frais.

Numéro sans frais pour les **demandes de renseignements** (Canada et États-Unis) : 1 800 263-1136
Appareils de télécommunications pour les **malentendants** : 1 800 363-7629
Numéro sans frais pour **commander seulement** (Canada et États-Unis) : 1 800 267-6677
Numéro sans frais de commande **par télécopieur** (Canada et États-Unis) : 1 800 287-4369

Liste des centres régionaux de consultation de Statistique Canada

Région de l'Atlantique
Dessert Terre-Neuve et le Labrador, la Nouvelle-Écosse, l'Île-du-Prince-Édouard et le Nouveau-Brunswick
Statistique Canada
Services-conseils
1741, rue Brunswick
2e étage, boîte 11
Halifax (Nouvelle-Écosse) B3J 3X8
Appels sans frais : 1 800 263-1136
Appels locaux : (902) 426-5331
Télécopieur : (902) 426-9538
Courrier électronique : atlantic.info@statcan.ca

Région du Québec
Dessert tout le Québec, sauf la région de la Capitale nationale et le territoire du Nunavut
Statistique Canada
Services-conseils
200, boul. René-Lévesque Ouest
Complexe Guy-Favreau
4e étage, Tour Est
Montréal (Québec) H2Z 1X4

Appels sans frais : 1 800 263-1136
Appels locaux : (514) 283-5725
Télécopieur : (514) 283-9350
Courrier électronique : louise.bournot@statcan.ca

Région de la Capitale nationale
Dessert la région de la Capitale nationale
Statistique Canada
Centre de consultation statistique
Immeuble R.-H.-Coats, Rez-de-chaussée
Avenue Holland
Ottawa (Ontario) K1A 0T6
Appels sans frais : 1 800 263-1136
Appels locaux : (613) 951-8116
Télécopieur : (613) 951-0581
Courrier électronique : infostats@statcan.ca
Si vous demeurez à l'extérieur de la zone de communication locale, composez le numéro sans frais de votre région.

Région de l'Ontario
Dessert tout l'Ontario, sauf la région de la Capitale nationale
Statistique Canada
Services-conseils
Immeuble Arthur-Meighen, 10e étage
25, avenue St. Clair Est
Toronto (Ontario) M4T 1M4
Appels sans frais : 1 800 263-1136
Appels locaux : (416) 973-6586
Télécopieur : (416) 973-7475

Région des Prairies
Dessert le Manitoba, la Saskatchewan, l'Alberta et les Territoires du Nord-Ouest

Pour le Manitoba
Statistique Canada
Services-conseils
Immeuble Via Rail, pièce 200
123, rue Main
Winnipeg (Manitoba) R3C 4V9
Appels sans frais : 1 800 263-1136
Appels locaux : (204) 983-4020
Télécopieur : (204) 983-7543
Courrier électronique : statswpg@solutions.net

Pour la Saskatchewan :
Statistique Canada
Services-conseils
Park Plaza, suite 440
2365, rue Albert
Regina (Saskatchewan) S4P 4K1

Appels sans frais : 1 800 263-1136
Appels locaux : (306) 780-5405
Télécopieur : (306) 780-5403
Courrier électronique : statcan@sk.sympatico.ca

Pour le nord de l'Alberta et les Territoires du Nord-Ouest :
Statistique Canada
Services-conseils
Park Square, 15e étage
10001, Bellamy Hill
Edmonton (Alberta) T5J 3B6
Appels sans frais : 1 800 263-1136
Appels locaux : (780) 495-3027
Télécopieur : (780) 495-5318
Courrier électronique : ewieall@statcan.ca

Pour le sud de l'Alberta :
Statistique Canada
Services-conseils
Discovery Place, pièce 201
3553, 31e rue N.-O.
Calgary (Alberta) T2L 2K7
Appels sans frais : 1 800 263-1136
Appels locaux : (403) 292-6717
Télécopieur : (403) 292-4958
Courrier électronique : degagnej@cadvision.com

Région du Pacifique
Dessert la Colombie-Britannique et le Territoire du Yukon
Statistique Canada
Services-conseils
Library Square Office Tower
600 - 300, rue Georgia Ouest
Vancouver (Colombie-Britannique) V6B 6C7
Appels sans frais : 1 800 263-1136
Appels locaux : (604) 666-3691
Télécopieur : (604) 666-4863
Courrier électronique : stcvan@statcan.ca